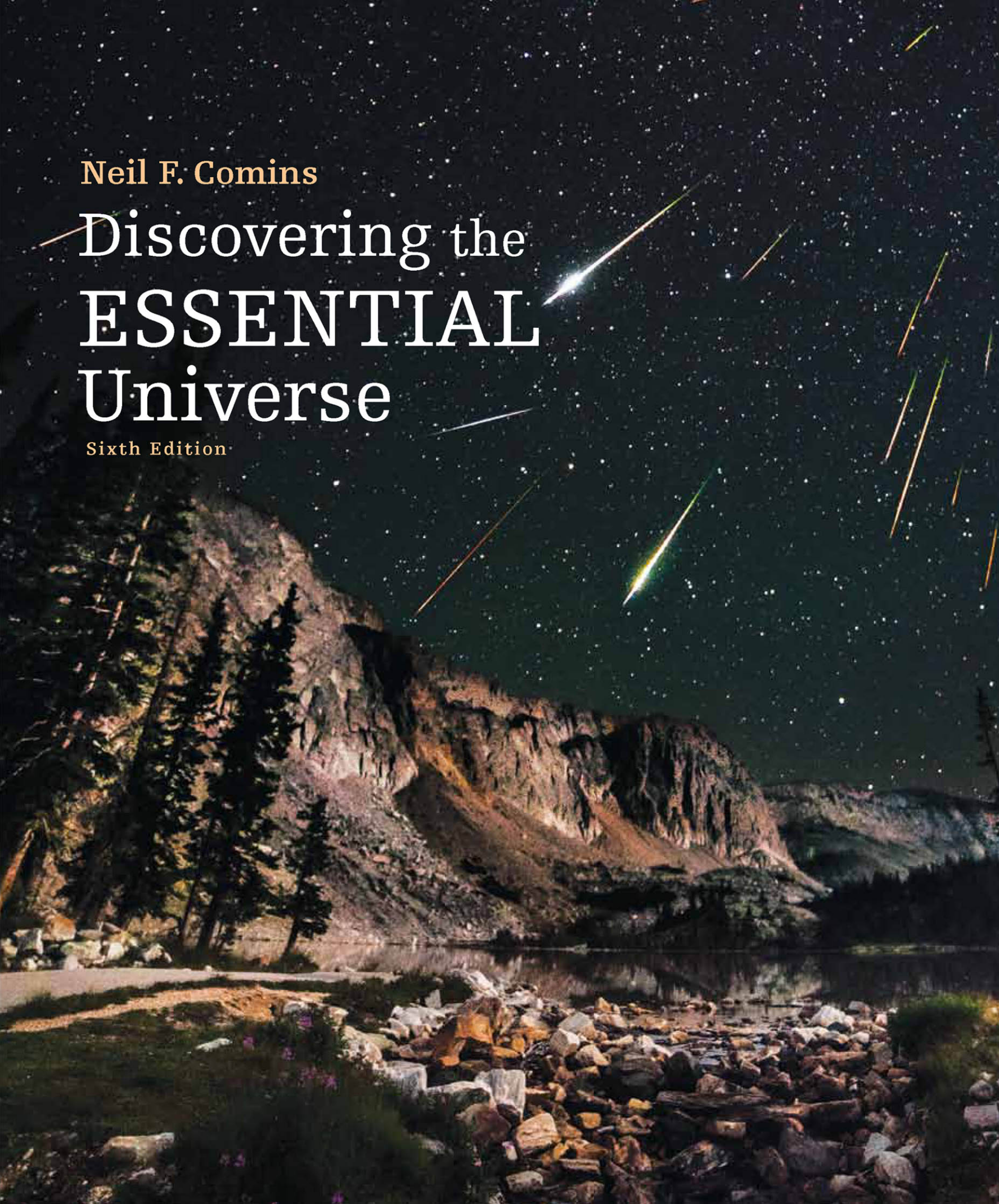


A night sky filled with stars and several bright meteor streaks in orange, yellow, and green. Below the sky is a rugged mountain landscape with a rocky foreground, a calm lake reflecting the sky, and a forest of evergreen trees on the left.

Neil F. Comins

Discovering the ESSENTIAL Universe

Sixth Edition

This page intentionally left blank

Neil F. Comins
University of Maine

Discovering the ESSENTIAL Universe

Sixth Edition

 W. H. FREEMAN
& COMPANY

A Macmillan Education Imprint

Dedicated to Edith Shapiro, Lenny Harris, and to the
memory of Bernard Shapiro.

Publisher: Katherine Parker
Acquisitions Editor: Alicia Brady
Development Editor: Brittany Murphy
Project Manager: Sofia Buono, codeMantra
Senior Media and Supplements Editor: Amy Thorne
Marketing Manager: Taryn Burns
Marketing Assistant: Bailey James
Editorial Assistant: Nandini Ahuja
Cover and Interior Design: Victoria Tomaselli
Director of Editing, Design, and Media Production: Tracey Kuehn
Production Coordinator: Paul Rohloff
Managing Editor: Lisa Kinne
Art Manager: Matthew McAdams
Photo Editor: Sheena Goldstein
Photo Researcher: Nicholas Ciani
Composition: codeMantra
Printing and Binding: RR Donnelley
Cover Photo: © 2012 David Kingham

Library of Congress Preassigned Control Number: 2014950361

ISBN-13: 978-1-4641-8170-2

ISBN-10: 1-4641-8170-5

© 2015, 2013, 2009, 2006 by W. H. Freeman and Company

All rights reserved

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke
RG21 6XS, England
www.whfreeman.com

CONTENTS OVERVIEW

Preface	xi	CHAPTER 9 The Sun: Our Extraordinary Ordinary Star	267
About the Author	xix	CHAPTER 10 Characterizing Stars	293
CHAPTER 1 Discovering the Night Sky	1	CHAPTER 11 The Lives of Stars from Birth Through Middle Age	317
CHAPTER 2 Gravitation and the Motion of the Planets	27	CHAPTER 12 The Deaths and Remnants of Stars	347
CHAPTER 3 Light and Telescopes	49	CHAPTER 13 The Galaxies	389
CHAPTER 4 Formation of the Solar System	97	CHAPTER 14 Cosmology	443
CHAPTER 5 Exoplanets	113	CHAPTER 15 Astrobiology	469
CHAPTER 6 The Terrestrial Planets and Their Moons	125	APPENDICES	A-1
CHAPTER 7 The Outer Planets and Their Moons	191	Glossary	G-1
CHAPTER 8 Vagabonds of the Solar System	233	Answers to Computational Questions	ANS-1
		Index	I-1
		EULA	

This page intentionally left blank

CONTENTS

Preface	xi		
About the Author	xix		
CHAPTER 1			
Discovering the Night Sky	1		
NAVIGATING THE NIGHT SKY	2		
1-1 The night sky is full of patterns	2		
1-2 Constellations make locating stars easy	3		
1-3 The celestial sphere aids in navigating the sky	5		
1-4 An “alt”ernative coordinate system	6		
1-5 Earth orbits the Sun in a plane called the ecliptic	6		
EARTHLY CYCLES	8		
1-6 Earth’s rotation creates the day-night cycle and its revolution defines a year	8		
THE SEASONS	10		
1-7 The seasons result from the tilt of Earth’s rotation axis combined with Earth’s revolution around the Sun	10		
1-8 Clocks and calendars are based on Earth’s rotation and revolution	13		
1-9 Precession is a slow, circular motion of Earth’s axis of rotation	14		
THE PHASES OF THE MOON	16		
1-10 The phases of the Moon originally inspired the concept of the month	16		
ECLIPSES	18		
1-11 Eclipses do not occur during every new or full Moon phase	18		
1-12 Three types of lunar eclipses occur	19		
1-13 Three types of solar eclipses also occur	20		
SCALES OF THE UNIVERSE	23		
1-14 Astronomical distances are, well, astronomical	23		
Summary of Key Ideas	25		
CHAPTER 2			
Gravitation and the Motion of the Planets	27		
SCIENCE: KEY TO COMPREHENDING THE COSMOS	28		
2-1 Science is both a body of knowledge and a process of learning about nature	28		
		CHANGING OUR EARTH-CENTERED VIEW OF THE UNIVERSE	30
		2-2 The belief in a Sun-centered cosmology formed slowly	30
		DISCOVERY 2-1 Earth-Centered Universe	32
		2-3 Copernicus devised the first comprehensive heliocentric cosmology	32
		2-4 Tycho Brahe made astronomical observations that disproved ancient ideas about the heavens	36
		KEPLER’S AND NEWTON’S LAWS	37
		2-5 Kepler’s laws describe orbital shapes, changing speeds, and the lengths of planetary years	37
		DISCOVERY 2-2 Units of Astronomical Distance	40
		2-6 Galileo’s discoveries strongly supported a heliocentric cosmology	41
		2-7 Newton formulated three laws that describe fundamental properties of physical reality	42
		2-8 Newton’s description of gravity accounts for Kepler’s laws	44
		Summary of Key Ideas	46
		CHAPTER 3	
		Light and Telescopes	49
		THE NATURE OF LIGHT	50
		3-1 Newton discovered that white is not a fundamental color and proposed that light is composed of particles	50
		3-2 Light travels at a finite but incredibly fast speed	53
		3-3 Einstein showed that light sometimes behaves as particles that carry energy	54
		3-4 Visible light is only one type of electromagnetic radiation	55
		OPTICS AND TELESCOPES	57
		3-5 Reflecting telescopes use mirrors to concentrate incoming starlight	57
		3-6 Telescopes brighten, resolve, and magnify	60
		3-7 Eyepieces, refracting telescopes, and binoculars use lenses to focus incoming light	62
		3-8 Shaping telescope mirrors and lenses is an evolving science	64
		3-9 Storing and analyzing light from space is key to understanding the cosmos	66
		3-10 Earth’s atmosphere hinders astronomical research	68

CAPTURING NONVISIBLE LIGHT: NONOPTICAL ASTRONOMY 71

- 3-11 Specially designed telescopes gather electromagnetic energy in all of the nonvisible parts of the spectrum 71

BLACKBODY RADIATION 79

- 3-12 An object's peak color shifts to shorter wavelengths as it is heated 79
- 3-13 The relative intensities of different emitted colors reveal a star's surface temperature 82

IDENTIFYING THE ELEMENTS BY ANALYZING THEIR UNIQUE SPECTRA 84

- 3-14 Each chemical element produces its own unique set of spectral lines 84
- 3-15 The various brightness levels of spectral lines depend on conditions in the spectrum's source 87

ATOMS AND SPECTRA 88

- 3-16 An atom consists of a small, dense nucleus surrounded by electrons 88
- 3-17 Spectra occur because electrons absorb and emit photons with only certain wavelengths 90
- 3-18 Spectra provide information about motion of objects toward or away from us but not across the sky 92

Summary of Key Ideas 94

CHAPTER 4

Formation of the Solar System 97

THE SOLAR SYSTEM CONTAINS HEAVY ELEMENTS, FORMED FROM AN EARLIER GENERATION OF STARS 98

- 4-1 Stars transform matter from lighter elements into heavier ones 98
- 4-2 Gravity, rotation, collisions, and heat shaped the young solar system 99

THE FORMATION OF THE PLANETS 103

- 4-3 The giant planets formed in sequence 103
- 4-4 The inner planets formed primarily from collisions 104

DEBRIS: REMNANTS IN THE SOLAR SYSTEM 105

- 4-5 The changing orbits of the giant planets spread debris throughout the solar system 105
- 4-6 The asteroid belt is leftover debris 107
- 4-7 The infalling debris from the giant planets led to the Late Heavy Bombardment 107

CATEGORIES OF THE PRESENT-DAY SOLAR SYSTEM 108

- 4-8 The categories of solar system objects have evolved 108
- 4-9 The orbits of the planets are related 109
- 4-10 The Sun developed while the planets matured 110
- Summary of Key Ideas 110

CHAPTER 5

Exoplanets 113

EXOPLANETS—PLANETS OUTSIDE OUR SOLAR SYSTEM 114

- 5-1 Protoplanetary disks are a common part of the star-forming process 114
- 5-2 Astronomers have at least seven different ways of detecting planets outside our solar system 114
- 5-3 Exoplanets orbit a breathtaking variety of stars 118
- 5-4 Exoplanets with a wide range of sizes, masses, and compositions have been observed 118
- 5-5 Stars with multiple planets have been observed 120
- 5-6 Many exoplanets have extraordinary orbits, as compared to those in our solar system 120
- 5-7 Planets that are not orbiting stars have also been observed 120
- 5-8 There are billions and billions of planets 122
- 5-9 Planets with liquid water are being discovered 122
- 5-10 The search for life on exoplanets is under way 123
- Summary of Key Ideas 123

CHAPTER 6

The Terrestrial Planets and Their Moons 125

COMPARATIVE PLANETOLOGY 126

- 6-1 Comparisons of the eight planets show distinct similarities and significant differences 126

EARTH: A DYNAMIC, VITAL WORLD 129

- 6-2 Earth's atmosphere has evolved over billions of years 130
- 6-3 Plate tectonics produce major changes on Earth's surface 133
- 6-4 Earth's interior consists of a rocky mantle and an iron-rich core 135
- 6-5 Earth's magnetic field shields us from the solar wind 137

THE MOON AND TIDES 141

- 6-6 The Moon's surface is covered with craters, plains, and mountains 141
- 6-7 Visits to the Moon yielded invaluable information about its history 145
- 6-8 The Moon probably formed from debris cast into space when a huge planetesimal struck the young Earth 150
- 6-9 Tides have played several important roles in the history of Earth and the Moon 152
- 6-10 The Moon is moving away from Earth 154

MERCURY 154

- 6-11 Photographs from *Mariner 10* and *Messenger* spacecraft reveal Mercury's lunarlike surface 154
- 6-12 Mercury has a higher percentage of iron than Earth 158
- 6-13 Mercury's rotation and revolution are coupled 160

6-14 Mercury's atmosphere is the thinnest of all terrestrial planets	161
--	-----

VENUS 162

6-15 The surface of Venus is completely hidden beneath a permanent cloud cover	162
6-16 The greenhouse effect heats Venus's surface	164
6-17 Venus is covered with gently rolling hills, two "continents," and numerous volcanoes	165

MARS 168

6-18 Mars's global features include plains, canyons, craters, and volcanoes	169
6-19 Although no canals exist on Mars, it does have some curious natural features	172
6-20 Mars's interior is less molten than the inside of Earth	172
6-21 Martian air is thin and often filled with dust	175
6-22 Surface and underground features indicate that water once flowed on Mars	177
6-23 Search for microscopic life on Mars continues	182
6-24 Mars's two moons look more like potatoes than spheres	183

COMPARATIVE PLANETOLOGY OF THE INNER PLANETS 184

6-25 Comparisons of planetary features provide new insights	184
Summary of Key Ideas	186

CHAPTER 7

The Outer Planets and Their Moons 191

JUPITER 193

7-1 Jupiter's outer layer is a dynamic area of storms and turbulent gases	193
7-2 Jupiter's interior has four distinct regions	196
7-3 Impacts provide probes into Jupiter's atmosphere	198

JUPITER'S MOONS AND RINGS 199

7-4 Io's surface is sculpted by volcanic activity	201
7-5 Europa harbors liquid water below its surface	202
7-6 Ganymede is larger than Mercury	204
7-7 Callisto bears the scars of a huge asteroid impact	205
7-8 Other debris orbits Jupiter as smaller moons and ringlets	205

SATURN 207

7-9 Saturn's atmosphere, surface, and interior are similar to Jupiter	207
7-10 Saturn's spectacular rings are composed of fragments of ice and ice-coated rock	210
7-11 Titan has a thick atmosphere, clouds, and lakes filled with liquids	215
7-12 Rhea has ice	218
7-13 Enceladus has water jets, an atmosphere, and a magnetic field	219

URANUS 220

7-14 Uranus sports a hazy atmosphere and clouds	220
7-15 A system of rings and satellites revolves around Uranus	222

NEPTUNE 224

7-16 Neptune was discovered because it had to be there	226
7-17 Neptune has rings and captured moons	226

COMPARATIVE PLANETOLOGY OF THE OUTER PLANETS 228

Summary of Key Ideas	230
----------------------	-----

CHAPTER 8

Vagabonds of the Solar System 233

DWARF PLANETS 234

8-1 Pluto and its moon, Charon, are about the same size	234
8-2 Ceres is a dwarf planet in the asteroid belt, while Pluto, Eris, Haumea, and Makemake are trans-Neptunian objects as well as dwarf planets	237

SMALL SOLAR SYSTEM BODIES 238

ASTEROIDS 238

8-3 Most asteroids orbit the Sun between Mars and Jupiter	238
8-4 Jupiter's gravity creates gaps in the asteroid belt	241
8-5 Asteroids also orbit outside the asteroid belt	244

COMETS 246

8-6 Comets come from far out in the solar system	246
8-7 Comet tails develop from gases and dust pushed outward by the Sun	249
8-8 Comets are fragile yet durable	252
8-9 Comets do not last forever	254

METEORIDS, METEORS, AND METEORITES 256

8-10 Small, rocky debris peppers the solar system	256
8-11 Meteorites are space debris that land intact	258
8-12 The Allende meteorite provides evidence of catastrophic explosions	262
8-13 Asteroid impacts with Earth have caused mass extinctions	263

Summary of Key Ideas	264
----------------------	-----

CHAPTER 9

The Sun: Our Extraordinary Ordinary Star 267

THE SUN'S ATMOSPHERE 269

9-1 The photosphere is the visible layer of the Sun	269
9-2 The chromosphere is characterized by spikes of gas called spicules	269
9-3 The outermost layer of the Sun's atmosphere, the corona, is exceptionally hot	271

THE ACTIVE SUN 273

- 9-4 Sunspots reveal the solar cycle and the Sun's rotation 273
- 9-5 The Sun's magnetic fields create sunspots 276
- 9-6 Solar magnetic fields also create other atmospheric phenomena 279

THE SUN'S INTERIOR 282

- 9-7 Thermonuclear reactions in the core of the Sun produce its energy 282
- 9-8 The solar model describes how energy escapes from the Sun's core 283

DISCOVERY 9-1 Thermonuclear Fusion 284

- 9-9 The Sun has gotten brighter over time 287
- 9-10 The mystery of the missing neutrinos inspired research into the fundamental nature of matter 287
- Summary of Key Ideas** 289

CHAPTER 10**Characterizing Stars** 293**LEAVING THE SOLAR SYSTEM** 294

- 10-1 Distances to nearby stars are found using stellar parallax 294

DISCOVERY 10-1 Distances to Nearby Stars 296**MAGNITUDE SCALES** 296

- 10-2 Apparent magnitude measures the brightness of stars as seen from Earth 297
- 10-3 Absolute magnitudes and luminosities do not depend on distance 298

DISCOVERY 10-2 The Distance–Magnitude Relationship 299**THE TEMPERATURES OF STARS** 299

- 10-4 A star's color reveals its surface temperature 300
- 10-5 A star's spectrum also reveals its surface temperature 301
- 10-6 Stars are classified by their spectra 302

TYPES OF STARS 303

- 10-7 The Hertzsprung-Russell diagram identifies distinct groups of stars 303
- 10-8 Luminosity classes set the stage for understanding stellar evolution 305
- 10-9 A star's spectral type and luminosity class provide a second distance-measuring technique 306

DISCOVERY 10-3 Kepler's Third Law and Stellar Masses 307**STELLAR MASSES** 307

- 10-10 Binary stars provide information about stellar masses 307
- 10-11 Main-sequence stars have a relationship between mass and luminosity 310
- 10-12 The orbital motion of binary stars affects the wavelengths of their spectral lines 311

Summary of Key Ideas 313**CHAPTER 11****The Lives of Stars from Birth Through Middle Age** 317**PROTOSTARS AND PRE-MAIN-SEQUENCE STARS** 318

- 11-1 Gas and dust exist between the stars 318
- 11-2 Supernovae, collisions of interstellar clouds, and starlight trigger new star formation 322
- 11-3 When a protostar ceases to accumulate mass, it becomes a pre-main-sequence star 324
- 11-4 The evolutionary track of a pre-main-sequence star depends on its mass 324
- 11-5 H II regions harbor young star clusters 326
- 11-6 Plotting a star cluster on an H-R diagram reveals its age 328

MAIN-SEQUENCE AND GIANT STARS 330

- 11-7 Stars spend most of their lives on the main sequence 330

EVOLUTION OF LOW MASS (0.08–0.4 $M_{\odot}$) STARS 332

- 11-8 Red dwarfs convert essentially their entire mass into helium 332

EARLY AND MIDDLE EVOLUTION OF INTERMEDIATE (0.4–8 $M_{\odot}$) AND HIGH-MASS STARS 333

- 11-9 When core hydrogen fusion slows down, a main-sequence star with mass greater than 0.4 $M_{\odot}$ becomes a giant 333
- 11-10 Helium fusion begins at the center of a giant 334
- 11-11 Life in the giant phase has its ups and downs 335

VARIABLE STARS 336

- 11-12 A Cepheid pulsates because it is alternately expanding and contracting 337
- 11-13 Cepheids enable astronomers to estimate vast distances 337
- 11-14 Globular clusters are bound groups of old stars 338
- 11-15 Mass transfer in close binary systems can produce unusual double stars 340

Summary of Key Ideas 344**CHAPTER 12****The Deaths and Remnants of Stars** 347**INTERMEDIATE-MASS (0.4 $M_{\odot}$ –8 $M_{\odot}$) STARS AND PLANETARY NEBULAE** 348

- 12-1 Intermediate-mass stars become supergiants before expanding into planetary nebulae 349
- 12-2 The burned-out core of an intermediate-mass star becomes a white dwarf 351

HIGH-MASS STARS (GREATER THAN 8 $M_{\odot}$) AND TYPE II SUPERNOVAE 354

- 12-3 A series of fusion reactions in high-mass stars leads to luminous supergiants 354

12-4	High-mass stars blow apart in Type II supernova explosions	355
12-5	Supernova remnants are observed in many places	358
NEUTRON STARS AND PULSARS		362
12-6	The cores of many Type II supernovae become neutron stars	362
12-7	A rotating magnetic field explains the pulses from a neutron star	364
12-8	Neutron stars have internal structure	366
12-9	Colliding neutron stars may provide some of the heavy elements in the universe	367
12-10	Binary neutron stars create pulsating X-ray sources	367
THE RELATIVITY THEORIES		369
12-11	Einstein revolutionized our understanding of space, time, and gravity	369
INSIDE A BLACK HOLE		374
12-12	Matter in a black hole becomes much simpler than elsewhere in the universe	374
12-13	Falling into a black hole is an infinite voyage	376
EVIDENCE FOR BLACK HOLES		377
12-14	Several binary star systems contain black holes	377
12-15	Other black holes range in mass up to billions of solar masses	378
12-16	Black holes and neutron stars in binary systems often create jets of gas	380
GAMMA-RAY BURSTS		381
12-17	Gamma-ray bursts are the most powerful explosions in the known universe	381
12-18	Black holes evaporate	383
Summary of Key Ideas		385

CHAPTER 13

The Galaxies		389
THE MILKY WAY		391
13-1	Studies of Cepheid variable stars revealed that the Milky Way is only one of many galaxies	391
13-2	Cepheid variables help us locate our Galaxy's center	392
13-3	Nonvisible observations help map the galactic disk	394
13-4	The galactic nucleus is an active, crowded place	397
13-5	Our Galaxy's disk is surrounded by a two-shell spherical halo of stars and other matter	399
13-6	The Galaxy is rotating	400
MYSTERIES AT THE GALACTIC FRINGES		402
13-7	Most of the matter in the Galaxy has not yet been identified	402

TYPES OF GALAXIES		403
13-8	The winding of a spiral galaxy's arms is correlated to the size of its central bulge	403
13-9	Explosions create flocculent spirals, and waves create grand-design spirals	406
13-10	Bars of stars run through the central bulges of barred spiral galaxies, and some disk galaxies, the lenticulars, lack spiral arms	408
13-11	Elliptical galaxies display a wide variety of sizes and masses	409
13-12	Galaxies without global structure are called irregular	411
13-13	Hubble presented spiral and elliptical galaxies in a tuning fork-shaped diagram	411
CLUSTERS AND SUPERCLUSTERS		412
13-14	Galaxies exist in clusters that may form still larger clumps called superclusters	412
13-15	Galaxies in a cluster can collide and combine	416
13-16	Dark matter helps hold together individual galaxies and clusters of galaxies	419
SUPERCLUSTERS IN MOTION		422
13-17	The redshifts of superclusters indicate that the universe is indeed expanding	422
DISCOVERY 13-1 The Tully-Fisher Relation and Other Distance-Measuring Techniques		424
DISCOVERY 13-2 The Expanding Universe		425
13-18	Astronomers are looking back to a time when galaxies were first forming	425
QUASARS		426
13-19	Quasars look like stars but have huge redshifts	427
OTHER ACTIVE GALAXIES		430
13-20	Active galaxies can be either spiral or elliptical	430
SUPERMASSIVE ENGINES		433
13-21	Supermassive black holes exist at the centers of most galaxies	433
13-22	Jets of protons and electrons ejected from around black holes help explain active galaxies	435
13-23	Gravity focuses light from quasars	437
Summary of Key Ideas		438

CHAPTER 14

Cosmology		443
COSMOLOGY		444
14-1	General relativity predicts an expanding (or contracting) universe	444
14-2	The expansion of the universe creates a Dopplerlike redshift	445
14-3	The Hubble constant is related to the age of the universe	445

THE BIG BANG	446	15-3 Evidence is mounting that life might exist elsewhere in our solar system	474
14-4 Remnants of the Big Bang have been detected	446	15-4 Searches for advanced civilizations try to detect their radio signals	474
14-5 The universe has two symmetries—isotropy and homogeneity	448	15-5 The Drake equation: How many civilizations are likely to exist in the Milky Way?	477
A BRIEF HISTORY OF SPACETIME, MATTER, ENERGY, AND EVERYTHING	449	15-6 Humans have been sending signals into space for more than a century	478
14-6 All physical forces in nature were initially unified	449	Summary of Key Ideas	478
14-7 Equations explain the evolution of the universe, even before matter and energy, as we know them, existed	450	APPENDICES	
14-8 Homogeneity and isotropy are results of inflation	451	A Powers-of-Ten Notation	A-1
14-9 During the first second, most of the matter and antimatter in the universe annihilated each other	452	B Temperature Scales	A-1
14-10 The universe changed from being controlled by radiation to being controlled by matter	453	C The Planets: Orbital Data	A-3
THE STRUCTURE OF THE UNIVERSE	455	D The Planets: Physical Data	A-3
14-11 Galaxies formed from huge clouds of primordial gas	455	E Major Satellites of the Planets by Mass	A-4
14-12 Star formation activity determines a galaxy's initial structure	459	F The Nearest Stars	A-5
THE FATE OF THE UNIVERSE	460	G The Visually Brightest Stars	A-6
14-13 The average density of matter is one factor that determines the future of the universe	460	H The Constellations	A-7
14-14 The overall shape of spacetime affects the future of the universe	461	I Some Useful Astronomical Quantities	A-10
14-15 Dark energy is causing the universe to accelerate outward	463	J Some Useful Physical Constants	A-10
Summary of Key Ideas	465	K Common Conversions between U.S. Customary and Metric Units	A-10
DISCOVERY 14-1 Superstring Theory and M-Theory	466	L Mass and Energy Inventory for the Universe	A-11
CHAPTER 15		M Reading Graphs	A-11
Astrobiology	469	N Periodic Table of the Elements	A-14
15-1 Astrobiology connects the cosmos and the origins of life	470	O Tides	A-15
15-2 The existence of life depends on chemical and physical properties of matter	471	P Energy and Momentum	A-16
		Q Radioactivity and the Ages of Objects	A-18
		R Gravitational Force	A-18
		S Largest Optical Telescopes in the World	A-19
		Glossary	G-1
		Answers to Computational Questions	ANS-1
		Index	I-1
		EULA	

PREFACE

Discovering the Essential Universe, Sixth Edition, is one of the briefest texts available for the introductory astronomy course, while still providing the wide range of topics that is a hallmark of the Comins/Kaufmann texts. A variety of ways to learn and to teach are provided for in this text and in the ancillary material that accompanies it. For example, questions for in-class interactivities, where students respond to questions with “clickers,” are provided to help enrich the classroom experience. Online materials provide tutorials and practice questions that turn students from passive into active learners.

The sixth edition of *Discovering the Essential Universe* continues to present concepts clearly and accurately, while strengthening the pedagogical tools to make the learning process even more worthwhile by

- presenting the underlying physical concepts needed to connect astronomical observations to the theories that explain them coherently and meaningfully.
- using both textual and graphical information to present concepts for students who learn in different ways.
- addressing student misconceptions in a respectful but rigorous manner, helping readers to understand why modern scientific views are correct.
- using analogies from everyday life to make cosmic phenomena more concrete.

NEW TO THIS EDITION OF *DISCOVERING THE ESSENTIAL UNIVERSE*

Exoplanets The explosion of exoplanet discoveries has prompted us to add a chapter (Chapter 5) on this subject. This chapter directly follows the chapter on the formation of the solar system, reinforcing student understanding of how our own planetary system formed while showing how others also have formed.

Coverage of the altazimuth coordinate system helps students to visualize their position in the cosmos (Chapter 1).

Revised coverage of planet classification The categories of *planets*, *dwarf planets*, and *small solar system bodies* are explained and reconciled with the existing classes of objects, including planets, moons, asteroids, meteoroids, and comets. Also explained is how Pluto fits more comfortably with the dwarf planets than with the eight planets in our solar system.

Expanded coverage of the satellites of the planets in our solar system is included.

Many scientific updates are included in this edition, with new findings in many areas of physics and astronomy, including the discovery of the Higgs boson, indications

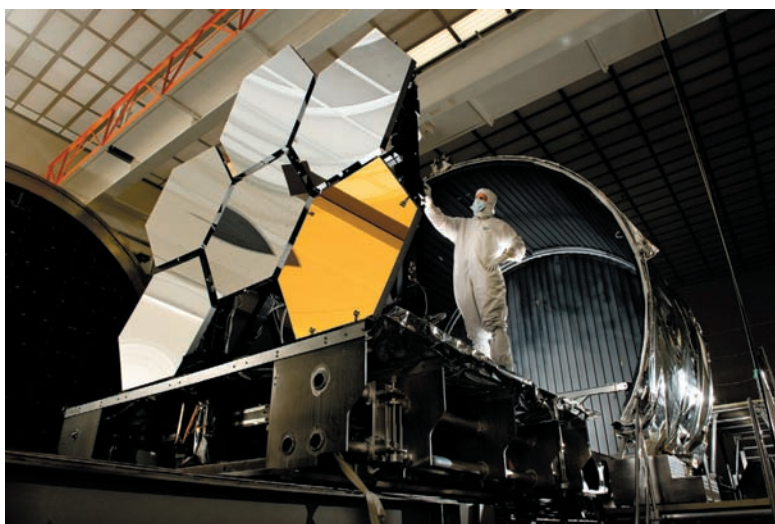


FIGURE 5-10 Aerospace lead optical test engineer Dave Chaney inspects six of the 18 primary mirror segments of NASA's James Webb Space Telescope, prior to cryogenic (extremely low temperature) testing in the X-ray & Cryogenic Facility at NASA's Marshall Space Flight Center in Huntsville,

Alabama. The James Webb Space Telescope will be launched in 2018 to study the formation of the first stars and galaxies and shed new light on the evolution of the universe. (NASA/MSFC/David Higginbotham)

of water activity on Mars, observations of fragmentation of an asteroid and rings around another one, the meeting of the *Rosetta* spacecraft with Comet 67P/Churyumov-Gerasimenko, the discovery of new exoplanets with interesting orbits and masses, the unusual sunspot cycle we are in, the improved parallax measurements by *Gaia* and by the Hubble Space Telescope, and new properties of the solar wind, among many other things.

PROVEN FEATURES SUPPORT LEARNING

What Do You Think? and **What Did You Think?** questions in each chapter ask students to consider their present beliefs and actively compare them with the correct science presented in the book. Numbered icons mark the places in the text where each concept is discussed. Encouraging students to think about what they believe is true and then working through the correct science step-by-step has proved to be an effective teaching technique, especially when time constraints prevent instructors from working with students individually or in small groups as they try to reconcile incorrect beliefs with proper science.

Focus Question 2-6

We saw in Chapter 1 that the Moon's orbit around Earth is not circular. Where in its orbit is the Moon moving fastest, and where is it moving slowly?

Focus Questions about important concepts are presented in most sections of the book. These questions encourage students to frequently test themselves and correct their beliefs before errors accumulate. For example, after learning about inertia in Section 2-7, students are asked how, while driving in a car, they can show that their bodies have inertia. (For instance, they could put on the brakes and feel themselves being restrained by the seatbelts.) **Focus Questions** are now numbered and placed in the margins for easy assignability.

Dynamic art Summary figures appear throughout the book to show either the interactions between important concepts or the evolution of important objects. For example, the location of the Sun in the sky, which varies over the seasons, as do the corresponding intensity of the light and the appropriate ground cover, is shown in a sequence of drawings combined into one figure.

Learning objectives underscore the key chapter concepts.

Section headings are brief sentences that summarize section content and serve as a quick study guide to the chapter when reread.

WHAT DO YOU THINK?

- 1 Can Earth's ozone layer, which has been partially depleted, be naturally replenished?
- 2 Do we see all parts of the Moon's surface at some time throughout the lunar cycle?

WHAT DID YOU THINK?

- 1 *Can Earth's ozone layer, which has been partially depleted, be naturally replenished?* Yes. Ozone is created continuously from normal oxygen molecules by their interaction with the Sun's ultraviolet radiation.
- 2 *Do we see all parts of the Moon's surface at some time during the lunar cycle?* No. Because the Moon's rotation around Earth is synchronous, we always see the same side. The far side of the Moon has been seen only from spacecraft that pass or orbit it.

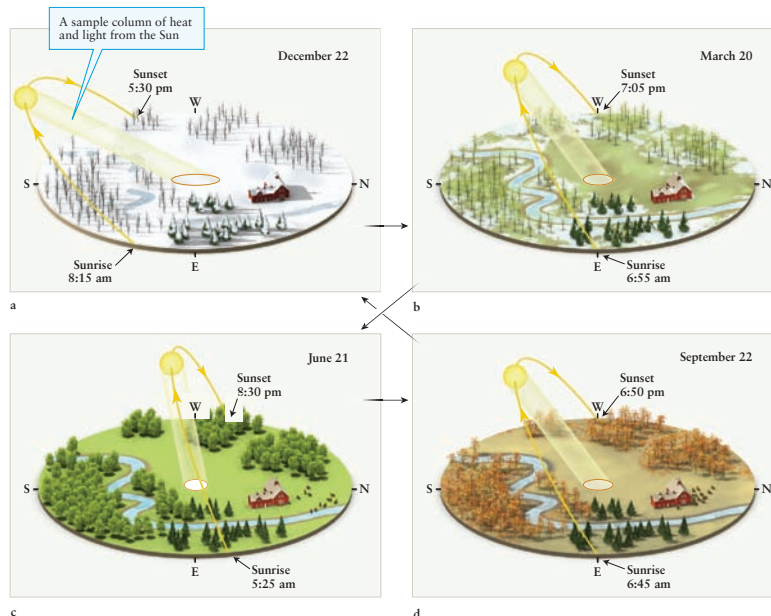


FIGURE 1-16 The Sun's Daily Path and the Energy It Deposits Here Northern Hemisphere: (a) On the winter solstice—the first day of winter—the Sun rises farthest south of east, is lowest in the noontime sky, stays up the shortest time, and its light and heat are least intense (most spread out) of any day of the year in the northern hemisphere. (b) On the vernal equinox—the first day of spring—the Sun rises precisely in the east and sets precisely in the west. Its light and heat have been growing more intense, as shown by the brighter oval of light than in (a). (c) On the summer solstice—the first day of summer—the Sun rises farthest north of east of any day in the year, is highest in the noontime sky, stays up the longest

time, and its light and heat are most intense of any day in the northern hemisphere. (d) On the autumnal equinox, the same astronomical conditions exist as on the vernal equinox. **Southern Hemisphere:** If you are reading this in the southern hemisphere, make the following changes: (a) Change December 22 to June 21 and visualize the Sun's path starting and ending the same distance *north* of east and north of west as it is south of east and south of west as shown here; (b) change March 20 to September 23; (c) change June 21 to December 21 and visualize the Sun's path starting and ending the same distance *south* of east and south of west as shown north of east and north of west here; (d) change September 22 to March 20.

DISCOVERY 2-1

Earth-Centered Universe

Most of us find it hard to understand why anyone would believe that the Sun, planets, and stars orbit Earth. After all, we *know* that Earth spins on its axis. We *know* that the gravitational force from the Sun holds the planets in orbit, just as Earth's gravitational force holds the Moon in orbit. In childhood these facts become part of our understanding of the motions of the heavenly bodies.

Psychologists call this background information that we use to help explain things a *conceptual framework*. A conceptual framework contains all of the information we take for granted. For example, when the Sun rises, moves across the sky, and sets, we take for granted that it is Earth's rotation that causes the Sun's apparent motion.

than those held in the past. Studying science helps us to develop intuition that is consistent with the actual workings of nature.

Geocentric Explanation of the Planets' Retrograde Motion The early Greeks, working in what we now know is an incorrect geocentric (Earth-centered) cosmology, developed many theories to account for the occasional retrograde motion of the planets and the resulting loops that the planets trace out against the background stars. One of the most successful (albeit incorrect) ideas was expounded by the last of the great ancient Greek astronomers, Ptolemy, who lived in Alexandria, Egypt. His basic concepts are sketched in **Figure D2-1**. Each

Discovery boxes provide deeper coverage of certain topics. In previous editions, some of this material was found only in the appendices but, in response to instructors using the text, key concepts have been pulled into the chapter where they appear in context but still remain as optional reading assignments.

Insight Into Science boxes are brief asides that relate topics to the nature of scientific inquiry and encourage critical thinking.

Wavelength tabs with photographic images show whether an image was made with radio waves (R), infrared radiation (I), visible light (V), ultraviolet light (U), X rays (X), or gamma rays (G).

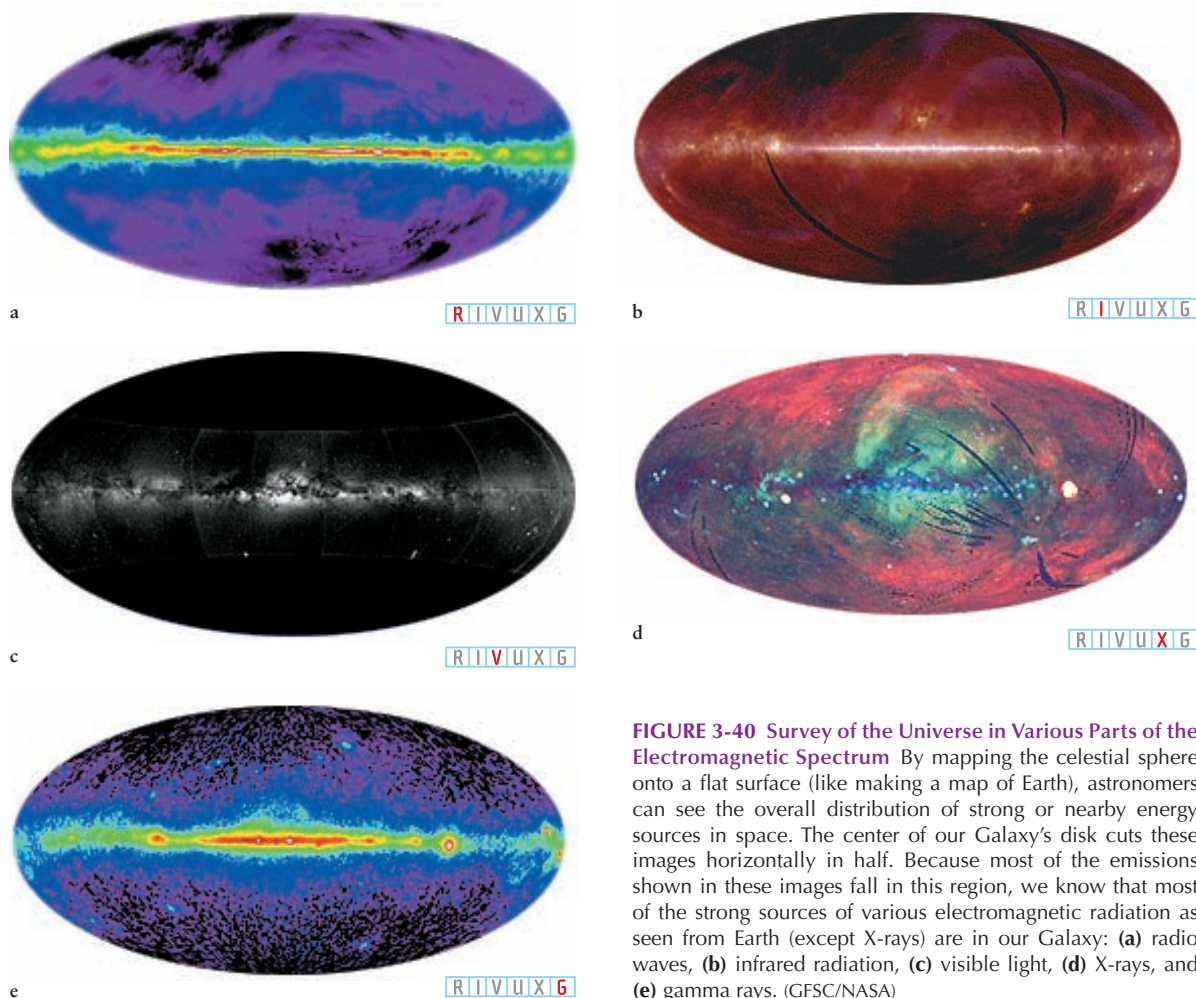


FIGURE 3-40 Survey of the Universe in Various Parts of the Electromagnetic Spectrum By mapping the celestial sphere onto a flat surface (like making a map of Earth), astronomers can see the overall distribution of strong or nearby energy sources in space. The center of our Galaxy's disk cuts these images horizontally in half. Because most of the emissions shown in these images fall in this region, we know that most of the strong sources of various electromagnetic radiation as seen from Earth (except X-rays) are in our Galaxy: (a) radio waves, (b) infrared radiation, (c) visible light, (d) X-rays, and (e) gamma rays. (GFSC/NASA)

Review and practice material

- **Summary of Key Ideas** is a bulleted list of key concepts.
- **What Did You Think?** questions at the end of each chapter answer the **What Do You Think?** questions posed at the beginning of the chapter.
- **Review Questions** and **Advanced Questions** help the students understand the chapter material.
- **Discussion Questions** offer interesting topics to spark lively, insightful debate.
- **NEW: Got It? Questions** ask students about either big picture concepts related to the chapter or questions associated with common misconceptions about that material (or both!).

MULTIMEDIA

FOR STUDENTS

Electronic Versions

Discovering the Essential Universe is offered in two electronic versions. One is an **Interactive e-Book**, available as part of LaunchPad, described below. The other is a PDF-based **e-Book from CourseSmart**, available through www.coursesmart.com. These options are provided to offer students and instructors flexibility in their use of course materials.

CourseSmart e-Book

Discovering the Essential Universe CourseSmart e-Book offers the complete text in an easy-to-use, flexible format. Students can choose to view the CourseSmart e-Book online or to download it to their computer or a mobile device, such as iPad, iPhone, or Android device. To help students study and mirror the experience of a printed textbook, CourseSmart e-Books feature note-taking, highlighting, and bookmark features.

ONLINE LEARNING OPTIONS

Discovering the Essential Universe supports instructors with a variety of online learning preferences. Its rich array of resources and platforms provides solutions according to each instructor's teaching method.

LaunchPad: Because Technology Should Never Get in the Way

At W. H. Freeman, we are committed to providing online instructional materials that meet the needs of instructors and students in powerful, yet simple ways—powerful enough to dramatically enhance teaching and learning, yet simple enough to use right away.

We have taken what we've learned from thousands of instructors and hundreds of thousands of students to create **LaunchPad**, a new generation of technology. LaunchPad offers our acclaimed content curated and organized for easy assignment in a breakthrough user interface in which power and simplicity go hand in hand.

Curated LaunchPad Units make class prep a whole lot easier

Combining a curated collection of video, tutorials, animations, projects, multimedia activities and exercises, and e-Book content, LaunchPad's interactive units give instructors a building block to use as is or as a starting point for their own learning units. An entire unit's worth of work can be assigned in seconds, drastically saving the amount of time it takes to have the course up and running.

LearningCurve

- Powerful adaptive quizzing, a game-like format, direct links to the e-Book, instant feedback, and the promise of better grades make using LearningCurve a no-brainer.
- Customized quizzing tailored to the text adapts to students' responses and provides material at different difficulty levels and topics based on student performance. Students love the simple yet powerful system and instructors can access class reports to help refine lecture content.

Chapter 2. Gravitation and the Motion of the Planets

In this unit, you will discover what makes a theory scientific. Copernicus revealed that Earth is not at the center of the universe, as previously believed, by observing the orbits of the planets. Subsequently, Kepler determined the shapes and other properties of planetary orbits by using the careful observations of his mentor Tycho Brahe. Finally, you will learn how Isaac Newton formulated an equation to describe the force of gravity and how he thereby explained why the planets and moons remain in orbit.

LearningCurve

Retrograde motion is an observed feature of the planets. It was explained in the Ptolemaic model of the solar system by having planets

- ☒ orbit the Earth in small circles moving on larger circles.
- ☐ orbit the Earth in nested circles, one within the other.
- ☐ move faster when close to the Earth and slower when farther away.
- ☐ orbit in a large circle but occasionally reverse their direction of travel.

Exactly! The correct answer is orbit the Earth in small circles moving on larger circles.

Take a Break Next Question

Interactive e-Book

The **Interactive e-Book** is a complete online version of the textbook with easy access to rich multimedia resources that complete student understanding. All text, graphics, tables, boxes, and end-of-chapter resources are included in the e-Book, and the e-Book provides instructors and students with powerful functionality to tailor their course resources to fit their needs.

- Quick, intuitive navigation to any section or subsection
- Full-text search, including the Glossary and Index
- Note-taking feature allows users to place notes at any highlighted text throughout the e-Book.
- “Top-note” feature allows users to place a prominent note at the top of the page to provide a more significant alert or reminder.
- Text highlighting, in a variety of colors

Astronomy Tutorials

- These self-guided, concept-driven experiential walk-throughs engage students in the process of scientific discovery as they make observations, draw conclusions, and apply their knowledge. Astronomy tutorials combine multimedia resources, activities, and quizzes.

Image Map Activities

- These activities use figures and photographs from the text to assess key ideas, helping students to develop their visual literacy. Students must click the appropriate section(s) of the image and answer corresponding questions.

Animations, Videos, Interactive Exercises, Flashcards

- Other LaunchPad resources highlight key concepts in introductory astronomy.

Assignments for Online Quizzing, Homework, and Self-Study

- Instructors can create and assign automatically graded homework and quizzes from the complete test bank, which is preloaded in LaunchPad. All quiz results feed directly into the instructor’s gradebook.

Scientific American Newsfeed

- To demonstrate the continued process of science and the exciting new developments in astronomy, the *Scientific American* Newsfeed delivers regularly updated material from the well-known magazine. Articles, podcasts, news briefs, and videos on subjects related to astronomy are selected for inclusion by *Scientific American*’s editors. The newsfeed provides several updates per week, and instructors can archive or assign the content they find most valuable.

Gradebook

- The included **gradebook** quickly and easily allows instructors to look up performance **metrics** for a whole class, for individual students, and for individual assignments. Having ready access to this information is help-

ful both in lecture prep and in making office hours more productive and efficient.

FOR INSTRUCTORS

Computerized Test Bank

Thousands of multiple-choice questions are section-referenced. The easy-to-use test bank allows instructors to add, edit, re-sequence, and print questions to suit their needs. The test bank can be downloaded from macmillanhigheered.com.

Online Course Materials (Blackboard, Moodle, Desire2Learn, Canvas)

As a service for adopters, we will provide content files in the appropriate online course format, including instructor and student resources for the textbook. The files can be used as is or can be customized to fit specific needs. Prebuilt quizzes, animations, and activities, among other materials, are included.

Classroom Presentation and Interactivity

A set of online lecture presentations created in PowerPoint allows instructors to tailor their lectures to suit their own needs using images and notes from the textbook. These presentations are available for download at macmillanhigheered.com.

Clicker Questions

Written by Neil Comins, these questions can be used as lecture launchers with or without a classroom response system such as iClicker. Each chapter includes questions relating to common misconceptions, as well as writing questions for instructors who would like to add a writing or class discussion element to their lectures.

FROM THE AUTHOR

“For me, it is far better to grasp the Universe as it really is than to persist in delusion, however satisfying and reassuring.”

— Carl Sagan, *The Demon-Haunted World: Science as a Candle in the Dark*

In 1888 the American astronomer Simon Newcomb wrote, “*We are probably nearing the limit of all we can know about astronomy.*” Boy, was he wrong. With the advent of photography in the nineteenth century and its continued refinement through CCDs and nonoptical telescope technology, combined with the development of the sciences in the twentieth century through today, and their engineering applications, our knowledge about the universe continues to expand at a phenomenal rate. New discoveries are being made in virtually every realm of astronomy: Every planet and moon we explore continue to reveal their secrets; we continue to discover new and fascinating properties of stars and stellar evolution; understanding the formation and evolution of galaxies,

like our Milky Way, is very much a work in progress; the properties of black holes, simple as these objects are, still beguile astrophysicists; the cause of the origin of the universe has yet to be determined; and the nature of 95% of the matter and energy in the universe is still unknown to science, among many other things.

To his credit, in 1903 Simon Newcomb acknowledged that “*What lies before us is an illimitable field.*” Today it is virtually impossible to keep up with all the new discoveries in astronomy, which is why astronomy textbooks such as this one must be updated so often. I have made every effort to provide your students with insights into the latest discoveries, along with the presentation of the underlying science and previous discoveries. I am also pleased to include a wide variety of modern learning techniques and new features in this sixth edition of *Discovering the Essential Universe* while still providing the wide range of factual topics that is the hallmark of the text.

In the realm of astronomy education, educators continue to develop methods to help students understand how to think like scientists and grasp the core concepts, even when scientific theories are at odds with students’ prior beliefs and misconceptions. I was a student at Cornell during Sagan’s hey-day there. It was memorable watching him blasting around in his orange and black Porsche 914. In the mid-1990s, just before his untimely death, he and I began discussing how to help students overcome their misconceptions about astronomy, an important theme of this book. Throughout each chapter, I have woven in explanations that will help your students identify, address, and hopefully overcome many of them.

ACKNOWLEDGMENTS

I am deeply grateful to the astronomers and teachers who reviewed the manuscript of this edition. This is a stronger and better book because of their conscientious efforts:

Robert Arts, *University of Pikeville*
 Sandi Brenner, *Bryant University*
 Sasa Dordevic, *University of Akron*
 Joseph Doyle, *Bridgewater State University*
 Adriana Durbala, *University of Wisconsin, Stevens Point*
 Robert Egler, *North Carolina State University*
 Brian Geislinger, *Gadsden State Community College*
 Bruce E. Gordon, *University of New Hampshire*
 George Hassel, *Siena College*
 Kishor T. Kapale, *Western Illinois University*
 Erik Kubik, *Quinnipiac University*
 Deborah A. Levine, *Glendale Community College*

Lauren Likkel, *University of Wisconsin, Eau Claire*
 Richard Ludlow, *Daniel Webster College*
 Jose Mena-Werth, *University of Nebraska, Kearney*
 Paul Meyer, *Lewis and Clark Community College*
 George Muncaster, *Scottsdale Community College*
 Chris Packham, *University of Texas at San Antonio*
 Marina Papenkova, *East Los Angeles College*
 Jon R. Pedicino, *College of the Redwoods*
 Chris Reese, *Lewis and Clark Community College*
 Joshua Ridley, *Murray State University*
 Aundrea Tavakkoly, *Porterville College*
 Victor M. Taveras, *West Kentucky Community and Technical College*
 Christopher Tycner, *Central Michigan University*
 Aaron Warren, *Purdue University, North Central*

I would also like to thank the many people whose advice on previous editions has had an ongoing influence:

John Anderson, *University of North Florida*
 Kurt S. J. Anderson, *New Mexico State University*
 Gordan Baird, *University of Arizona*
 Nadine G. Barlow, *Northern Arizona University*
 Henry E. Bass, *University of Mississippi*
 J. David Batchelor, *Community College of Southern Nevada*
 Jill Bechtold, *University of Arizona*
 Peter A. Becker, *George Mason University*
 Ralph L. Benbow, *Northern Illinois University*
 John Bieging, *University of Arizona*
 Greg Black, *University of Virginia*
 John C. (Jack) Brandt, *University of New Mexico*
 James S. Brooks, *Florida State University*
 John W. Burns, *Mt. San Antonio College*
 Gene Byrd, *University of Alabama*
 Eugene R. Capriotti, *Michigan State University*
 Eric D. Carlson, *Wake Forest University*
 Jennifer L. Cash, *South Carolina State University*
 Michael W. Castelaz, *Pisgah Astronomical Research Institute*
 Gerald Cecil, *University of North Carolina*
 David S. Chandler, *Porterville College*
 David Chernoff, *Cornell University*
 Erik N. Christensen, *South Florida Community College*
 Tom Christensen, *University of Colorado, Colorado Springs*
 Chris Clemens, *University of North Carolina*
 Christine Clement, *University of Toronto*
 Halden Cohn, *Indiana University*
 John Cowan, *University of Oklahoma*
 Antoinette Cowie, *University of Hawaii*
 Charles Curry, *University of Waterloo*

- James J. D’Amario, *Harford Community College*
Purnas Das, *Purdue University*
Peter Dawson, *Trent University*
John M. Dickey, *University of Minnesota, Twin Cities*
John D. Eggert, *Daytona Beach Community College*
Stephen S. Eikenberry, *University of Florida*
Bernd Enders, *College of Marin*
Rica French, *Mira Costa College*
Robert Frostick, *West Virginia State College*
Martin Gaskell, *University of Nebraska*
Richard E. Griffiths, *Carnegie Mellon University*
Bruce Gronich, *University of Texas, El Paso*
Siegbert Hagmann, *Kansas State University*
Thomasanna Hail, *Parkland College*
Javier Hasbun, *University of West Georgia*
David Hedin, *Northern Illinois University*
Chuck Higgins, *Penn State University*
Scott Hildreth, *Chabot College*
Thomas Hockey, *University of Northern Iowa*
Mark Hollabaugh, *Normandale Community College*
J. Christopher Hunt, *Prince George’s Community College*
James L. Hunt, *University of Guelph*
Nathan Israeloff, *Northeastern College*
Francine Jackson, *Framingham State College*
Kenneth Janes, *Boston University*
Fred Jacquin, *Onondaga Community College*
Katie Jore, *University of Wisconsin–Stevens Point*
William C. Keel, *University of Alabama*
William Keller, *St. Petersburg Junior College*
Marvin D. Kemple, *Indiana University–Purdue University Indianapolis (IUPUI)*
Pushpa Khare, *University of Illinois at Chicago*
Agnes Kim, *Georgia College & State University*
F. W. Kleinhaus, *Indiana University–Purdue University Indianapolis (IUPUI)*
Rob Klinger, *Parkland College*
H. S. La, *Clayton State University*
Patrick M. Len, *Cuesta College*
John Patrick Lestrade, *Mississippi State University*
C. L. Littler, *University of North Texas*
M. A. K. Lohdi, *Texas Tech University*
Michael C. LoPresto, *Henry Ford Community College*
Phyllis Luger, *Indiana University*
R. M. MacQueen, *Rhodes College*
Robert Manning, *Davidson College*
Paul Mason, *University of Texas, El Paso*
P. L. Matheson, *Salt Lake Community College*
Steve May, *Walla Walla Community College*
Rahul Mehta, *University of Central Arkansas*
Ken Menningen, *University of Wisconsin–Stevens Point*
J. Scott Miller, *University of Louisville*
Scott Miller, *Pennsylvania State University*
L. D. Mitchell, *Cambria County Area Community College*
J. Ward Moody, *Brigham Young University*
Siobahn M. Morgan, *University of Northern Iowa*
Steven Mutz, *Scottsdale Community College*
Charles Nelson, *Drake University*
Gerald H. Newsom, *Ohio State University*
Bob O’Connell, *College of the Redwoods*
William C. Oelfke, *Valencia Community College*
Richard P. Olenick, *University of Dallas*
John P. Oliver, *University of Florida*
Ron Olowin, *St. Mary’s College of California*
Melvyn Jay Oremland, *Pace University*
Jerome A. Orosz, *San Diego State University*
David Patton, *Trent University*
Jon Pedicino, *College of the Redwoods*
Sidney Perkowitz, *Emory University*
Lawrence Pinsky, *University of Houston*
Eric Preston, *Indiana State University*
David D. Reid, *Wayne State University*
Adam W. Rengstorf, *Indiana University*
James A. Roberts, *University of North Texas*
Henry Robinson, *Montgomery College*
Dwight P. Russell, *University of Texas, El Paso*
Barbara Ryden, *Ohio State University*
Itai Seggev, *University of Mississippi*
Larry Sessions, *Metropolitan State University*
C. Ian Short, *Florida Atlantic University*
Norm Siems, *Juniata College*
John D. Silva, *University of Massachusetts at Dartmouth*
Michael L. Sitko, *University of Cincinnati*
Earl F. Skelton, George F. Smoot, *University of California at Berkeley*
Alex G. Smith, *University of Florida*
Jack Sulentic, *University of Alabama*
David Sturm, *University of Maine, Orono*
Paula Szkody, *University of Washington*
Michael T. Vaughan, *Northeastern University*
Andreas Veh, *Kenai Peninsula College*
John Wallin, *George Mason University*
William F. Welsh, *San Diego State University*
R. M. Williamon, *Emory University*
Gerard Williger, *University of Louisville*
Wayne Wooten, *Pensacola Junior College*
Edward L. (Ned) Wright, *University of California at Los Angeles*
Jeff S. Wright, *Elon College*
Nicole E. B. Zellner, *Rensselaer Polytechnic Institute*

I would like to add my special thanks to the wonderfully supportive staff at W. H. Freeman and Company for making the revision process so enjoyable. Among others, these people include Kate Parker, Alicia Brady, Brittany Murphy, Paul Rohloff, Tracey Kuehn, Lisa Kinne, Amy Thorne, Victoria Tomaselli, Sheena Goldstein, Nick Ciani, and Matt McAdams. Thanks also go to Sofia Buono of codeMantra, my copyeditor Donna

Mulder, and to indexer Steve Katigbak. Warm thanks also to David Sturm, University of Maine, Orono, for his help in collecting current data on objects in the solar system; to my UM astronomy colleague David Batuski; and to my wife, Sue, for her patience and support while preparing this book.

Neil F. Comins
neil.comins@umit.maine.edu

ABOUT THE AUTHOR



Thomas Morelli Photography

Professor Neil F. Comins is on the faculty of the University of Maine. Born in 1951 in New York City, he grew up in New York and New England. He earned a bachelor's degree in engineering physics at Cornell University, a master's degree in physics at the University of Maryland, and a Ph.D. in astrophysics from University College, Cardiff, Wales, under the guidance of Bernard F. Schutz. Dr. Comins's work for his doctorate, on general relativity, was cited in Subramanyan Chandrasekhar's Nobel laureate speech. He has done theoretical and experimental research in general relativity, observational astronomy, computer simulations of galaxy evolution, and science education. The fourth edition of *Discovering the Universe* was the first book in this series that Dr. Comins wrote, having taken over following the death of Bill Kaufmann in 1994. He is also the author of four trade books, *What If the Moon Didn't Exist?*, *Heavenly Errors*, *The Hazards of Space Travel*, and *What If the Earth Had Two Moons?* His book *What If the Moon Didn't Exist?* has been made into planetarium shows, been excerpted for television and radio, translated into several languages, was the theme for the Mitsubishi Pavilion at the World Expo 2005 in Aichi, Japan, and has been a show at a resort since 2006. *Heavenly Errors* explores misconceptions people have about astronomy, why such misconceptions are so common, and how to correct them. Dr. Comins has appeared on numerous television and radio shows and gives many public talks. Although he has jumped out of airplanes while in the military, today his activities are a little more sedate: He is a licensed pilot and avid sailor, having once competed against Prince Philip, Duke of Edinburgh.

This page intentionally left blank



R I V I U X I G The central region of our Milky Way galaxy shines brightly above statues on Easter Island in the Pacific Ocean. Besides individual stars in the night sky, you can see clusters of stars, glowing clouds of gas, and darker regions of gas and dust obscuring light from stars behind them. We will explore all of these things and more in the chapters of this book. (Stephen Alvarez/ National Geographic/ Getty Images)

Discovering the Night Sky

WHAT DO YOU THINK?

- 1 What do astronomers define as constellations?
- 2 Is the North Star—Polaris—the brightest star in the night sky?
- 3 What causes the seasons?
- 4 When is Earth closest to the Sun?
- 5 How many zodiac constellations are there?
- 6 Does the Moon have a dark side that we never see from Earth?
- 7 Is the Moon ever visible during the daytime?
- 8 What causes lunar and solar eclipses?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

You are studying astronomy at an exciting time as astronomers draw open the curtains of the cosmos. Our understanding of the cosmos (or the *universe*) and how it evolves is growing as never before. That is due, in large measure, to the immense light-gathering power and sensitivity of modern telescopes. We can now see so far away and, therefore, so far into the past that we see the first stars and the first galaxies as they just began forming more than $13\frac{1}{2}$ billion years ago. We could not see these objects even a decade ago and, likewise, it took just 21 years for astronomers to discover 1000 planets orbiting nearby stars, a feat that would have been impossible 30 years ago.

In this chapter you will discover

- how astronomers map the night sky to help them locate objects in it
- that Earth's spin on its axis causes day and night
- how Earth's orbit around the Sun, combined with the tilt of Earth's axis of rotation relative to its orbit, creates the seasons
- that the Moon's orbit as seen from Earth creates the phases of the Moon and lunar and solar eclipses
- how the year is defined and how the calendar was developed

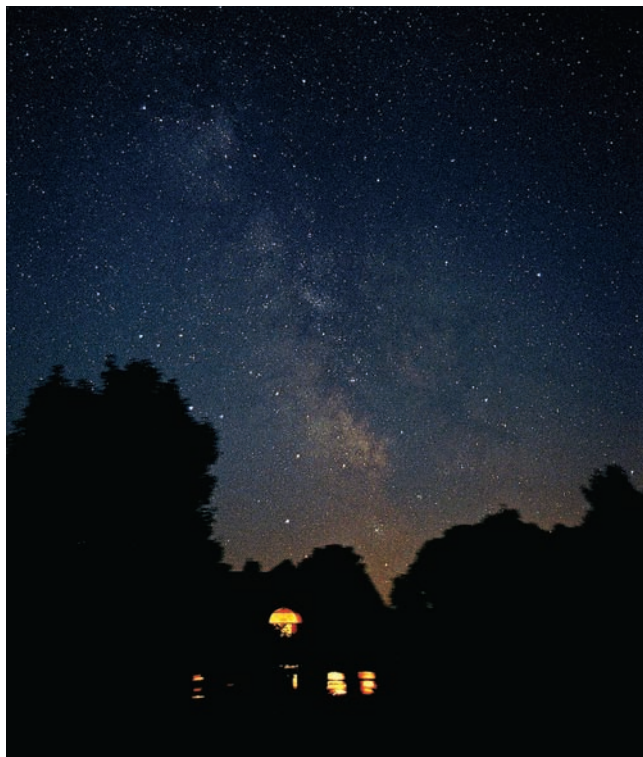
NAVIGATING THE NIGHT SKY

1-1 The night sky is full of patterns

When you gaze at the sky on a clear night, where the air is free of pollution and there is not too much light from cities or other sources, there seem to be millions of stars twinkling overhead. In reality, the unaided human eye can detect only about 6000 stars over the entire sky. At any one time, you can see roughly 3000 stars in

dark skies, because only half of the stars are above the *horizon*—the boundary between Earth and the sky. In very smoggy or light-polluted cities, you may see only a tenth of that number or less (**Figure 1-1**).

1 In any event, you probably have noticed patterns of bright stars, each formally called an **asterism**, and you are familiar with some common names for some of them, such as the ladle-shaped Big Dipper and broad-shouldered Orion. These recognizable patterns of stars



a

FIGURE 1-1 The Night Sky Without and With Light Pollution (a) Sunlight is a curtain that hides virtually everything behind it. As the Sun sets, places with little smog or light pollution treat viewers to beautiful panoramas of stars that can



b

RIVUXG

inspire the artist or scientist in many of us. This photograph shows the night sky in Goodwood, Ontario, Canada, during a power outage. (b) This photograph shows the same sky with normal city lighting. (© Todd Carlson/SkyNews Magazine)

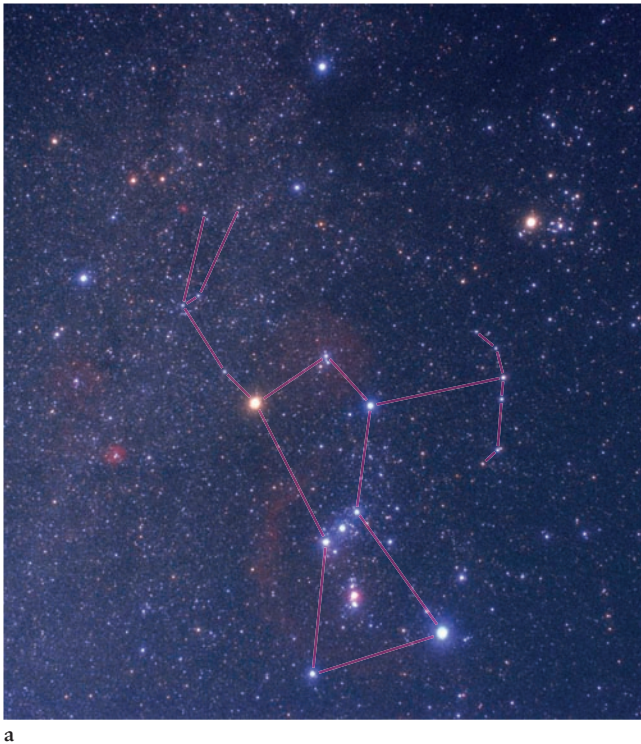
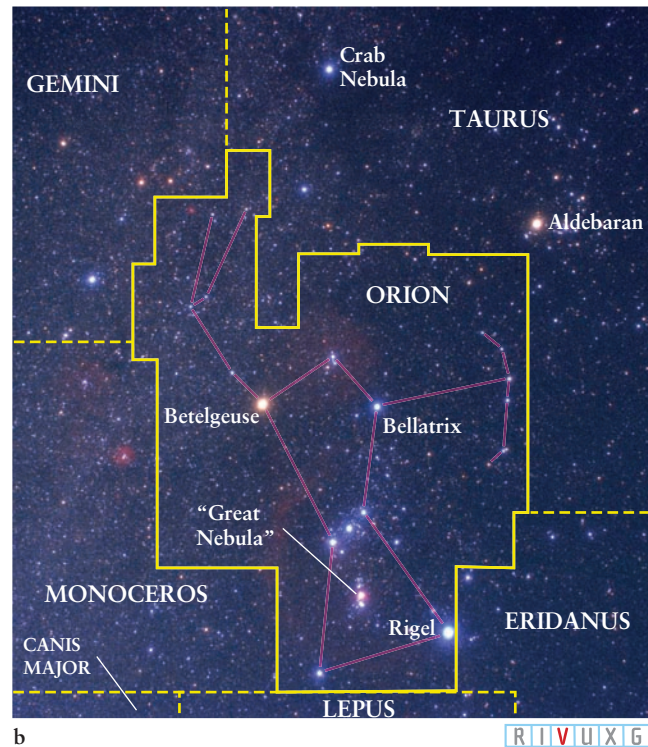


FIGURE 1-2 The Constellation Orion (a) The pattern of stars (*asterism*) called Orion is prominent in the winter sky. From the northern hemisphere, it is easily seen high above the southern horizon from December through March. You can see in this photograph that the various stars have different colors—something to watch for when you observe the night sky. (b) Technically, constellations are entire regions of



of the sky. The constellation called Orion and parts of other nearby constellations are depicted in this photograph. All the stars and other celestial objects, like galaxies, inside the boundary of Orion are members of that constellation. The entire sky is covered by a quiltlike pattern of 88 constellations of differing sizes and shapes. (© 2004 Jerry Lodriguss/Astropix.com)

(Figure 1-2a) are informally called constellations in everyday conversation. Technically, however, each **constellation** is an entire region of the sky and everything in that region. In what follows, we will often use the word “constellation” to mean either the asterisms or the regions of the sky. The context will tell you which version of the word is in use.

1-2 Constellations make locating stars easy

2 People have known for millennia how to find the direction north in locations where the Big Dipper is visible. To do this, locate the Big Dipper (the asterism in the constellation Ursa Major) and imagine that its bowl is resting on a table (Figure 1-3). If you see the dipper upside down in the sky, as you frequently will, imagine the dipper resting on an upside-down table above it. Locate the two stars of the bowl farthest from the Big Dipper’s handle. These are called the *pointer stars*. Draw a mental line through these stars in the direction away from the table, as shown in Figure 1-3. The first *moderately bright* star you then encounter is Polaris, also called the *North Star* because it is located almost directly over Earth’s North Pole. So, while Polaris is not

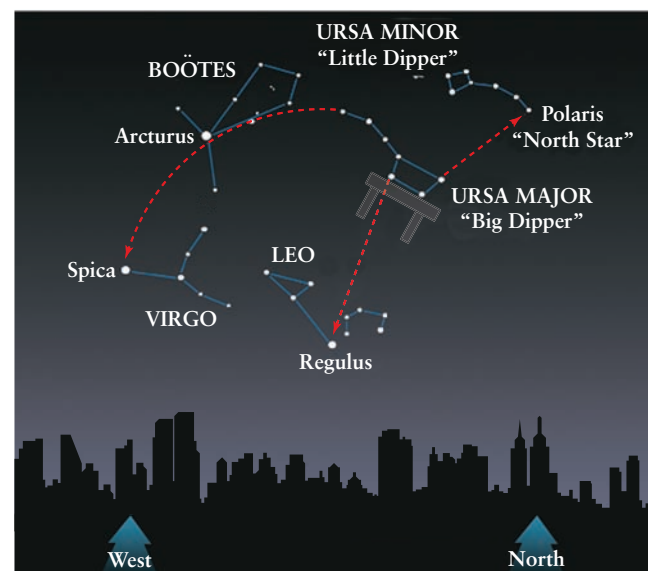


FIGURE 1-3 The Big Dipper as a Guide In the northern hemisphere, the Big Dipper is an easily recognized pattern of seven bright stars. This star chart shows how the Big Dipper can be used to locate Polaris (the North Star) as well as the brightest stars in three other constellations. While the Big Dipper appears right side up in this drawing of the sky shortly before sunrise, at other times of the night it appears upside down.

even among the 20 brightest stars (see Appendix G: The Visually Brightest Stars), it is easy to locate. Whenever you face Polaris, you are facing north. East is then on your right, south is behind you, and west is on your left. (There is no equivalent star over the South Pole.)

The Big Dipper also illustrates the fact that being familiar with just a few prominent constellations makes it easy to locate less distinctive constellations and their important stars. The most effective way to do this is to use vivid visual connections. For example, imagine gripping the handle of the Big Dipper and slamming its bowl straight down onto the head of Leo (the Lion). Leo comprises the first group of bright stars your dipper encounters. As shown in Figure 1-3, the brightest star in this group is Regulus, the dot of the backward question mark that traces the lion's mane. As another example, follow the arc of the Big Dipper's handle away from its bowl. The first bright star you encounter along that arc beyond the handle is Arcturus in Boötes (the

Herdsman). Follow the same arc farther to the prominent bluish star Spica in Virgo (the Virgin). Spotting these stars and remembering their names are easy if you remember the saying “Arc to Arcturus and speed on to Spica.”

Different constellations are visible at night during different times of the year. During the winter months in the northern hemisphere, you can see some

of the brightest stars in the sky. Many of them are in the

Focus Question 1-1

Referring to Figure 1-5, can you create a story to help you remember the connections between the constellations Cygnus (the Swan), Sagitta (the Arrow), and Aquila (the Eagle)?

Insight Into Science

Compare What You Believe with What Nature and Science Show You If you thought Polaris is the brightest star in the night sky, how did you learn that? There is likely to be a lot of information you learned from a variety of sources that you will have to unlearn while learning astronomy. While this process is often uncomfortable, it is essential if you want to understand how science works and how scientists view the world. Be careful: It is human nature to change new information to fit our current beliefs rather than change what we believe. Fight that urge if science can show you that your beliefs are incorrect!

vicinity of the “winter triangle,” which connects bright stars in the constellations of Orion (the Hunter), Canis Major (the Larger Dog), and Canis Minor (the Smaller Dog), as shown in Figure 1-4. The winter triangle passes high in the sky at night during the middle of winter. It is easy to find Sirius, the brightest star in the night sky, by locating the belt of Orion and following a straight mental line from it to the left (as you face Orion). The first bright star that you encounter is Sirius.

The “summer triangle,” which graces the summer sky as shown in Figure 1-5, connects the bright stars Vega in Lyra (the Lyre), Deneb in Cygnus (the Swan), and Altair in Aquila (the Eagle). A conspicuous portion

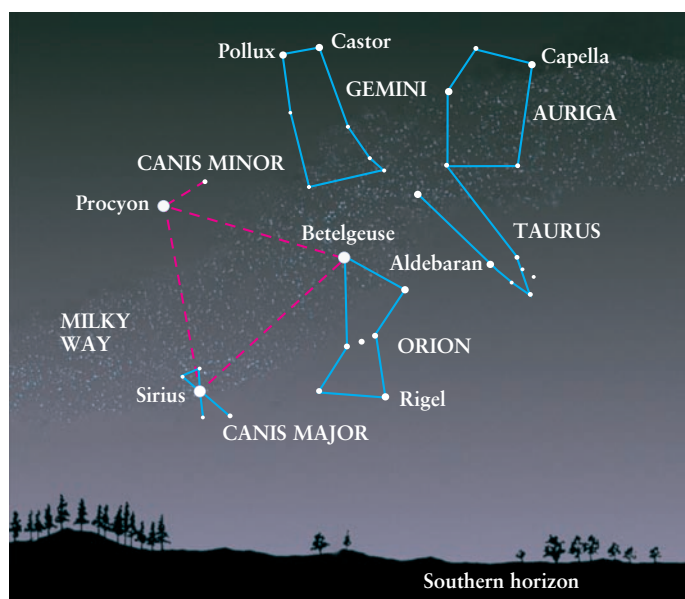


FIGURE 1-4 The Winter Triangle This star chart shows the southern sky in the northern hemisphere as it appears during the evening in December. Three of the brightest stars in the sky make up the winter triangle. In addition to the constellations involved in the triangle, Gemini (the Twins), Auriga (the Charioteer), and Taurus (the Bull) are also shown.

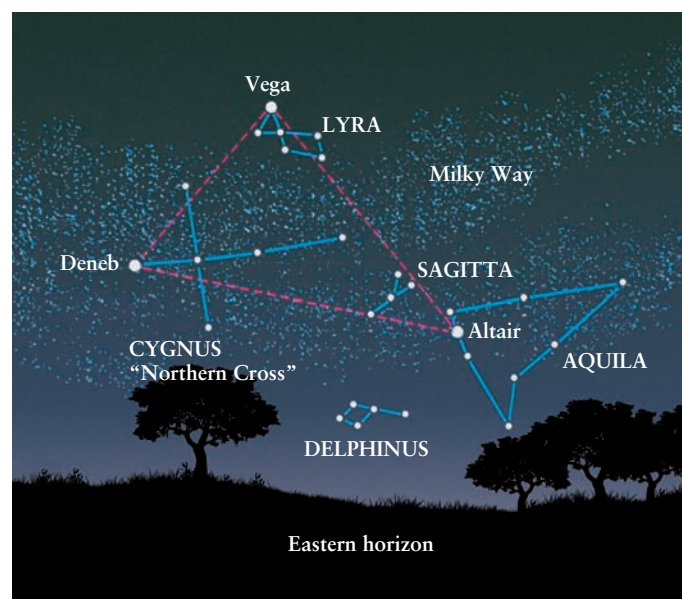


FIGURE 1-5 The Summer Triangle This star chart shows the northeastern sky as it appears in the evening in June. In addition to the three constellations involved in the summer triangle, the faint stars in the constellations Sagitta (the Arrow) and Delphinus (the Dolphin) are also shown.

of the Milky Way forms a beautiful background for these constellations, which are nearly overhead in the northern hemisphere during the middle of summer at midnight.

1-3 The celestial sphere aids in navigating the sky

Astronomers require more accuracy in locating dim objects than is possible simply by moving from asterism to asterism. They have therefore created a sky map, called the **celestial sphere**, and applied a coordinate system to it, analogous to the coordinate system of north-south latitude and east-west longitude used to navigate on Earth. If you know a star's celestial coordinates and know how to navigate around the night sky, you can locate it quickly. For such a sky map to be useful in finding stars, the stars must be fixed on it, just as cities are fixed on maps of Earth. If you look at the night sky year after year, you will see that the stars do indeed appear fixed relative to one another. Furthermore, throughout each night, entire patterns of stars appear to rigidly orbit Earth in fixed groups. We employ this artificial, Earth-based view of the heavens to make celestial maps by pretending that the stars are attached to the inside of an enormous hollow shell, the *celestial sphere*, with Earth at its center (**Figure 1-6a**). Visualized another way, you can imagine the half of the celestial sphere that is visible at night as a giant dome covering Earth (**Figure 1-6b**).

The 88 constellations that completely cover the sky are each formed around an asterism, which is why the constellations and their asterisms have the same names. These constellations were chosen in 1922 by the American astronomer Henry Norris Russell and their boundaries (see **Figure 1-2b**) were set in 1930 by the Belgian astronomer Eugene Delporte. This work was done under the auspices of the International Astronomical Union (IAU).

Some constellations, like Ursa Major (the Large Bear), are very large, while others, like Sagitta, are relatively small. To describe a star's location, we might say “Albireo in the constellation Cygnus,” much as we would refer to “Chicago in the state of Illinois,” “Melbourne in the state of Victoria,” or “Ottawa in the province of Ontario.”

The stars seem fixed on the celestial sphere only because of their remoteness. Stars are actually at a wide variety of distances from Earth, and they are moving relative to each other. But we neither see their motion nor perceive their relative distances because they are so far from us. You can understand this difficulty in seeing distant objects moving by first imagining a car driving right past you at 100 km/h (62 mi/h). Its motion is unmistakable. However, a plane traveling at 1000 km/h across the sky appears to be moving much more slowly. And an object at the distance of the Sun traveling at the same speed as the plane across the sky would appear to

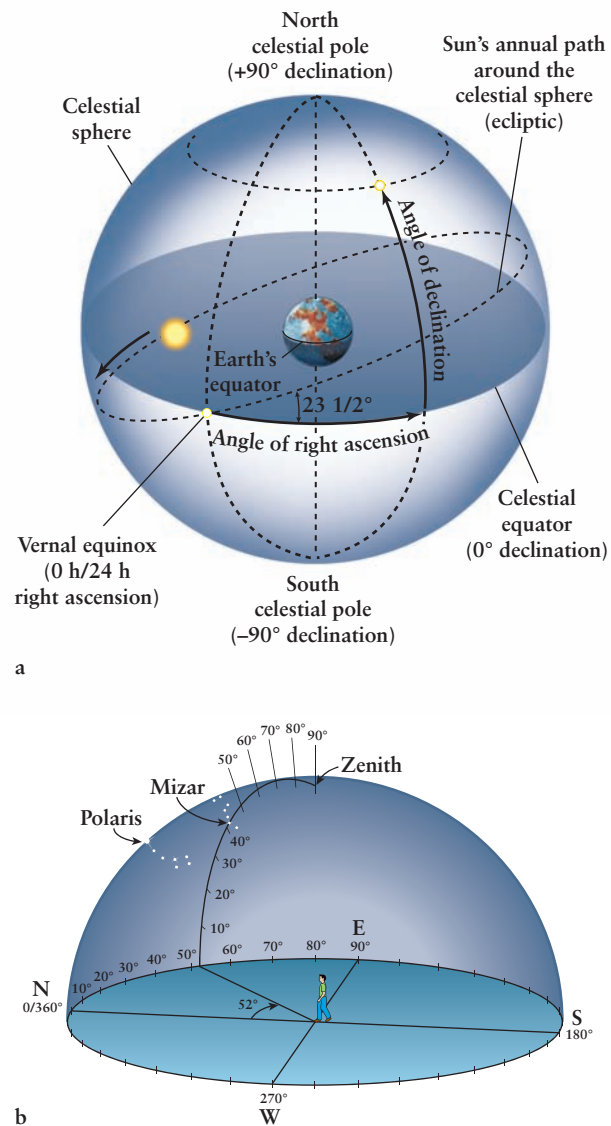


FIGURE 1-6 The Celestial Sphere and Coordinate Systems

(a) The celestial sphere is the apparent “bowl” or hollow sphere of the sky. The celestial equator and poles are projections of Earth’s equator and axis of rotation onto the celestial sphere. The north celestial pole is therefore located directly over Earth’s North Pole, while the south celestial pole is directly above Earth’s South Pole. Analogous to longitude and latitude on Earth, the coordinates in space are right ascension (R.A.) and declination (Dec.), respectively. The star in this figure has the indicated R.A. and Dec. (b) Altitude and azimuth create a local coordinate system for every place on Earth. Altitude is the angle straight up from the horizon. Azimuth is the angle eastward from north along the horizon. To find Mizar at 10 P.M. on March 14, 2013, as seen from Orono, Maine, we calculate its coordinates there as altitude 47° and azimuth 52° and use them to locate it in the night sky.

be going nearly 100 million times more slowly than that plane.

The stars (other than the Sun) are all more than 40×10^{12} km (40 trillion km or 25 trillion mi) from

us. Therefore, although the patterns of stars in the sky do change, their great distances prevent us from seeing those changes over the course of a human lifetime. Thus, as unrealistic as it is to assume that all the stars are fixed in space and that all are at the same distance from us, the celestial sphere is so useful for navigating the heavens that it is used by astronomers even at the most sophisticated observatories around the world.

As shown in Figure 1-6a, we can project key geographic features from Earth out into space to establish directions and bearings. If we expand Earth's equator onto the celestial sphere, we obtain the **celestial equator**, which divides the sky into northern and southern hemispheres, just as Earth's equator divides Earth into two hemispheres. We can also imagine projecting a straight line between Earth's North Pole and South Pole out into space above both poles. Where this line intersects the celestial sphere defines the **north celestial pole** and the **south celestial pole**, also shown in Figure 1-6a.

Using the celestial equator and poles as reference features, astronomers divide up the surface of the celestial sphere in precisely the same way that the latitude and longitude grid divides Earth. The equivalent to latitude on Earth is called **declination (Dec.)** on the celestial sphere. It is measured from 0° to 90° north or south of the celestial equator. The equivalent of longitude on Earth is called **right ascension (R.A.)** on the celestial sphere, measured in hours from 0 h to 24 h around the celestial equator (see Figure 1-6a). We derive the reference point (0 h) of right ascension, analogous to Earth's prime meridian, shortly. The boundaries of the constellations all meet at right angles and run along lines of constant right ascension and constant declination. Because stars are all so far away, their R.A. and Dec. do not change perceptibly over our lifetime (except as measured in modern observatories, which can detect changes for some of the nearer ones).

1-4 An "alt"ernative coordinate system

Just as latitude and longitude can take us any place on Earth, right ascension and declination are a coordinate system for the entire celestial sphere. If you know the R.A. and Dec. for a given object and know how to navigate around the celestial sphere (that is, if you know the angle for the object from the celestial equator and the angle around the celestial equator from the equivalent of the prime meridian on the sky), you can find that object's location in the night sky. The problem is that without lots of experience and/or the appropriate software at hand, figuring out the coordinate system on the night sky and locating that object can be difficult. There is an alternative, local coordinate system that often expedites navigating around the night sky, called the *alt-azimuth* or *altazimuth coordinate system*.

The alt-azimuth system (Figure 1-6b) depends on your location and the time of night. It is easy to use: First

find the North Pole star, Polaris (see Section 1-2). Facing the direction of Polaris, you are facing due north. Look straight at the horizon below Polaris. This is $(0^\circ, 0^\circ)$ in the alt-azimuth system. (Recall that $^\circ$ denotes the angular measure of degrees. There are 360° in a circle.) **Altitude** is the angle straight up from the horizon toward your **zenith**, the point directly overhead. (Every place has a different zenith.) Altitude runs between 0° at the horizon and 90° at the zenith. The angle to the right of north along the horizon is the **azimuth** coordinate for your position. For example, the azimuth of east is 90° , while the azimuth of south is 180° . Azimuth ranges from 0° to 360° . Figure 1-6b shows how to find Mizar, the middle star in the handle of the Big Dipper, given its altitude and azimuth at a certain location and time.

Because Earth spins on its axis, every object on the celestial sphere is continuously moving from our perspective, so while their R.A. and Dec. are constant, their altitude and azimuth angles are continually changing. Software like *Starry Night*TM can calculate the altitude and azimuth for any object, and because it is very easy to use the alt-azimuth system, as just described, these coordinates are frequently used in astronomy labs at schools and colleges.

1-5 Earth orbits the Sun in a plane called the ecliptic

Imagine that you could see the stars even during the day, so that you could follow the Sun's motion through the background constellations throughout the year. (The Sun appears to move among the stars because Earth orbits around it.) From day to day, you would see the Sun trace a straight path on the celestial sphere called the **ecliptic**. As you can see in **Figure 1-7a**, the ecliptic makes a closed circle bisecting the celestial sphere. Every year the Sun follows the same course through the sky because Earth follows the same path around it.

The term "ecliptic" has a second use in astronomy. Earth orbits the Sun in a plane also called the ecliptic (Figure 1-7b). Imagine yourself on the Sun watching Earth move day by day. The path of Earth on the celestial sphere as seen from the Sun is precisely the same as the path of the Sun as seen from Earth. In other words, the plane created by the Sun's path on the celestial sphere and the plane of Earth's orbit are identical. Hence, the two ecliptics are the same.

Note in both parts of this figure that Earth's axis of rotation, running through the North Pole and South Pole, is tilted relative to the ecliptic by $23\frac{1}{2}^\circ$. Because of this tilt, the celestial equator and the path of the Sun around the celestial sphere are tilted with respect to each other by that angle, as shown in Figure 1-6a. Because of this tilt, the two loops meet in two places. Where the Sun is heading northward, this intersection is called the **vernal equinox**. Halfway around the two circles they meet again at the **autumnal equinox**. At this

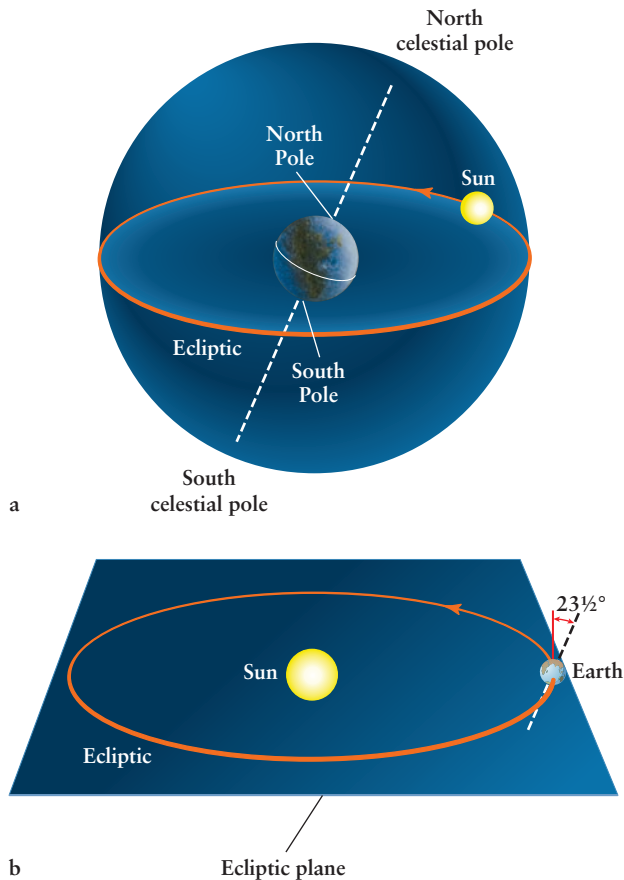


FIGURE 1-7 The Ecliptic (a) The ecliptic is the apparent annual path of the Sun on the celestial sphere. (b) The ecliptic is also the plane described by Earth's path around the Sun. The planes created by the two ecliptics exactly coincide. Note that the rotation axis through Earth's poles is tilted $23\frac{1}{2}^\circ$ from being perpendicular to the ecliptic.

time, the Sun is heading southward. Equinox comes from the Latin words meaning “equal night,” because when the Sun appears at either of these two points, it is directly over Earth's equator, resulting in 12 hours of daytime and 12 hours of nighttime everywhere on Earth on that day. Just as the location of Greenwich, England, defines the *prime meridian*, or zero of longitude on Earth, we need to establish a zero of right ascension. This zero value is defined to be at the vernal equinox (see Figure 1-6a).

In navigating around the celestial sphere, astronomers measure the relative positions of two objects by the angle between them. Ancient mathematicians invented a system of angles and angular measure that is still used today to denote the relative positions and apparent sizes of objects in the sky. To locate stars on the celestial sphere, for example, we do not need to know their distances from Earth. All we need to know is the angle from one star to another in the sky, a property that

remains fixed over our lifetimes because the stars are all so far away.

An **arc angle**, often just called an **angle**, is the opening between two lines that meet at a point. Angular measure is a method of describing the size of an angle. As noted earlier, the basic unit of angular measure is the **degree**, designated by the symbol $^\circ$. A full circle is divided into 360° . A right angle measures 90° . As shown in Figure 1-8, the angle between the two “pointer stars” in the Big Dipper is about 5° .

Astronomers also use angular measure to describe the apparent sizes of celestial objects. For example, imagine the full Moon. As seen from Earth, the angle across the Moon's diameter is nearly $\frac{1}{2}^\circ$. We therefore say that the **angular diameter**, or **angular size**, of the Moon is $\frac{1}{2}^\circ$. Alternatively, astronomers say that the Moon “subtends” an angle of $\frac{1}{2}^\circ$. In this context, subtend means “to extend across.”

To talk about smaller angles, we subdivide the degree into 60 arcminutes (abbreviated 60 arcmin or $60'$). An arcminute is further subdivided into 60 arcseconds (abbreviated 60 arcsec or $60''$). A dime viewed face-on from a distance of 1.6 km (1 mile) has an angular diameter of about $2''$. From everyday experience, we know that an object looks bigger when it is nearby than when it is far away. The angular size of an object therefore does not by itself tell you anything about its actual physical size. For example, the fact that the Moon's angular diameter is $\frac{1}{2}^\circ$ does not tell you how big the Moon really is. Nor does the Sun's angular size alone tell you how far it is from Earth.

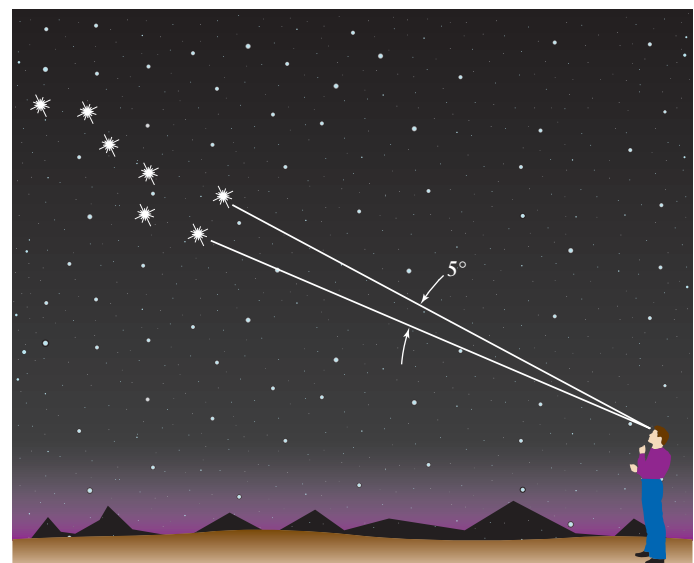


FIGURE 1-8 The Big Dipper The angular distance between the two “pointer stars” at the front of the Big Dipper is about 5° . For comparison, the angular diameter of the Moon is about $\frac{1}{2}^\circ$.

EARTHLY CYCLES

Earth spins on an axis running between its north and south poles, and it moves in an orbit around the Sun. We will now explore how these motions create the day-night cycle, the yearly cycle, and the seasons.

1-6 Earth's rotation creates the day-night cycle and its revolution defines a year

Rotation The spinning of a body on an axis running through it is called **rotation**. We do not feel Earth's rotation because our planet is so physically large compared to us that Earth's gravitational attraction holds us firmly on its surface. Earth's rotation causes the stars—as well as the Sun, the Moon, and the planets—to appear to rise along the eastern horizon, move across the sky, and set along the western horizon. Earth's daily rotation, causing the Sun to rise and set, thereby creates day and night. In general, this rotation creates **diurnal motion**, or daily motion, which causes all objects to circle around the celestial poles.

Depending on your latitude, some of the stars and constellations never disappear below the horizon. Instead, they trace complete circles in the sky over the course of each night (Figure 1-9). To understand why this happens, imagine that you are standing on Earth's North Pole at night. Looking straight up, you see Polaris. Because Earth is spinning around its axis directly under your feet, all the stars appear to move from left to right (counterclockwise) in horizontal rings above you. The exception is Polaris, which always remains at the North

Focus Question 1-2

Why are asterisms like the Big Dipper seen sometimes upright and sometimes upside down?



RIVUXIG

FIGURE 1-9 Circumpolar Star Trails Stars appear to rotate around Polaris, the North Star (small arc near point around which all arcs appear to orbit). Those stars that pass between Polaris and the horizon directly below it are circumpolar. The 4-m telescope dome at Kitt Peak National Observatory is directly below Polaris. The skylight is from Phoenix, Arizona, 160 km (100 mi) away. (Stan Honda/AFP/Getty Images)



FIGURE 1-10 Motion of Stars at the Poles Because Earth rotates around the axis through its poles, stars seen from these locations appear to move in huge, horizontal circles. This is the same effect you would get by standing up in a room and spinning around; everything would appear to move in circles around you. At the North Pole stars move left to right, while at the South Pole they move right to left.

Pole's zenith. As seen from the North Pole, no stars rise or set (Figure 1-10). They just seem to revolve around Polaris in horizontal circles. Stars and constellations that never go below the horizon as seen from a given place on Earth are called **circumpolar** (Figure 1-9). While there is no bright South Pole star equivalent to Polaris, all stars seen from the South Pole are also circumpolar and move from right to left (clockwise).

If you live in the northern hemisphere, Polaris is always located above your northern horizon at an angle equal to your latitude. Only the stars and constellations that pass between Polaris and the land directly below it are circumpolar. As you go farther south in the northern hemisphere, the number of stars and constellations that are circumpolar decreases. Likewise, as you go farther north in the southern hemisphere, the number of stars and constellations that are circumpolar also decreases.

Now visualize yourself at the equator. All the stars appear to rise straight up in the eastern sky and set straight down in the western sky (Figure 1-11). Polaris is barely visible on the northern horizon. While Polaris never sets, all the other stars do, and therefore none of the stars are circumpolar as seen from the equator.

As you can see from these last two mental exercises, the angle at which the stars rise and set depends on your viewing latitude. Figure 1-12 shows stars setting at 35° north latitude. Polaris is fixed at 35° above the horizon to the right of this figure, not at the zenith, as it is at the North Pole, nor on the horizon, as seen from the equator. As another example, except for those stars in the corners, all the stars whose paths are shown in Figure 1-9 are circumpolar because they are visible all night, every night.



FIGURE 1-11 Rising and Setting of Stars at the Equator Standing on the equator, you are perpendicular to the axis around which Earth rotates. As seen from there, the stars rise straight up on the eastern horizon and set straight down on the western horizon. This is the same effect you get when driving straight over the crest of a hill; the objects on the other side of the hill appear to move straight upward as you descend.

Revolution Whereas rotation is the spinning of an astronomical object around an axis through it, **revolution** is the motion of any astronomical object around another astronomical object. Earth takes 1 year, or about $365\frac{1}{4}$ days, to revolve around the Sun. A year on Earth is measured by the motion of our planet relative to the stars. For example, draw a straight line from the Sun through Earth to some star on the opposite side of Earth from the Sun. As Earth revolves around the Sun, that line inscribes a straight path on the celestial sphere (namely, the ecliptic) and returns to the original star $365\frac{1}{4}$ days later. The



R I V U X G

FIGURE 1-12 Rising and Setting of Stars at Middle North Latitudes Unlike the motion of the stars at the poles (see Figure 1-10), the stars at all other latitudes do change angle above the ground (that is, change astronomical “altitude”) throughout the night. This time-lapse photograph shows stars setting. The latitude determines the angle at which the stars rise and set. (David Miller/David Malin Images)

length of any cycle of motion measured with respect to the stars is called a **sidereal period**. Earth’s annual orbit around the Sun, 1 year, is a sidereal period, as is the period of the diurnal motion of the stars each day in our sky. This latter period is 23 h 56 m 4 s.

Different constellations are visible at night during different times of the year because of Earth’s revolution around the Sun. If Earth were rotating at a fixed place over the Sun, rather than orbiting around it, then every star would always rise and set at the same time every day. Of course, Earth is in motion around the Sun, and as a result the Sun moves around the celestial sphere and the stars rise approximately 4 minutes earlier each day than they did the day (or night) before. This effect accumulates, bringing different constellations up at night throughout the year. **Figure 1-13** summarizes

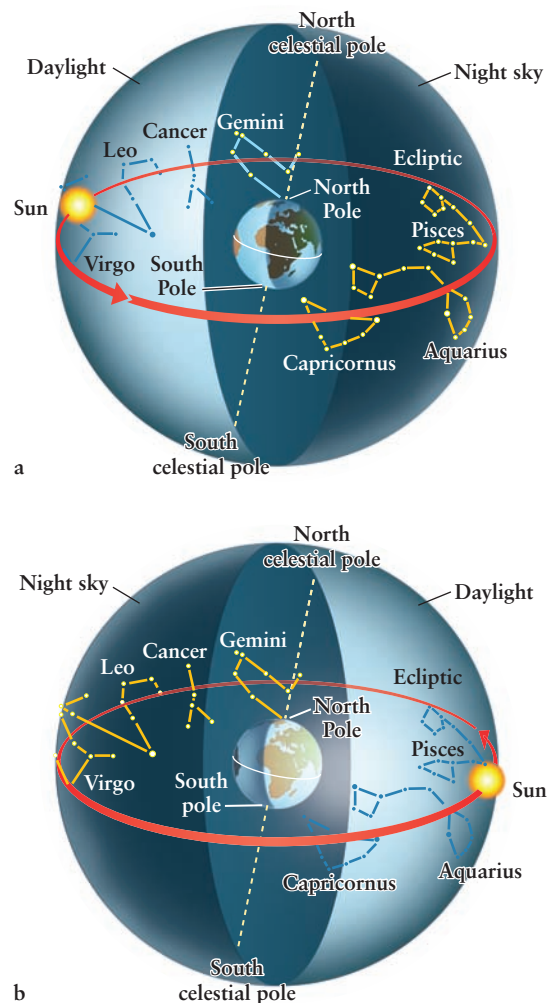


FIGURE 1-13 Why Different Constellations Are Visible at Different Times of the Year (a) On the autumnal equinox each year, the Sun is in the constellation Virgo. As seen from Earth, that part of the sky is in daylight, and we see stars only on the other half of the sky, centered around the constellation Pisces. (b) Six months later, the Sun is in Pisces. This side of the sky is then bright, while the side centered on Virgo is visible at night.

Insight Into Science

Define Your Terms As with interpersonal communication, using the correct words is very important in science. Scientific words usually have a specific meaning, such as “rotation” denoting spin and “revolution” meaning one object orbiting another. Be especially careful to understand the context in which words with more than one meaning, such as “ecliptic” and “constellation,” are used.

this motion. When the Sun is within the boundaries of (colloquially, “in”) Virgo (September 18–November 1), for example, the hemisphere containing the Sun and the constellations around Virgo are in daylight (see Figure 1-13a). When the Sun is up, so are Virgo and the surrounding constellations, so we cannot see them. During that time of year, the constellations on the other side of the celestial sphere, centered on the constellation Pisces, are in darkness. Thus, when the Sun is “in” Virgo, Pisces and the constellations around it are high in our sky at night.

Six months later, when the Sun is “in” Pisces, that half of the sky is filled with daylight, while Virgo and the constellations around it are high in the night sky (see Figure 1-13b). These arguments apply everywhere on Earth at the same time because the Sun moves along its path very slowly as seen from Earth, taking a year to make one complete circuit around the celestial sphere.

THE SEASONS

1-7 The seasons result from the tilt of Earth’s rotation axis combined with Earth’s revolution around the Sun

Equinoxes and solstices As we saw on Figure 1-6, the ecliptic and the celestial equator are different circles tilted $23\frac{1}{2}^\circ$ with respect to each other on the celestial sphere. This occurs because Earth’s rotation axis is tilted $23\frac{1}{2}^\circ$ away from a line perpendicular to the ecliptic (Figure 1-14 and see Figure 1-7). These two circles intersect at only two points, which are exactly opposite each other on the celestial sphere (Figure 1-15 and see Figure 1-6a). Each of these two points is called an **equinox**. Recall from Section 1-5 that on these two days, the Sun is directly over Earth’s equator, resulting in 12 hours of daytime and 12 hours of nighttime everywhere on Earth on that day.

Except for tiny changes each year, Earth maintains this tilted orientation as it orbits the Sun. Therefore, Polaris is above the North Pole throughout the year. For half the year, the northern hemisphere is tilted toward the Sun, and as a result, the Sun rises higher in the northern hemisphere’s sky than it does during the other half of the year (Figure 1-16). Equivalently, when the southern hemisphere is tilted toward the Sun, the Sun rises higher in the southern hemisphere’s sky.

Consider the location of the Sun throughout the year as seen from the northern hemisphere. The day that the Sun rises farthest south of east is around December 22

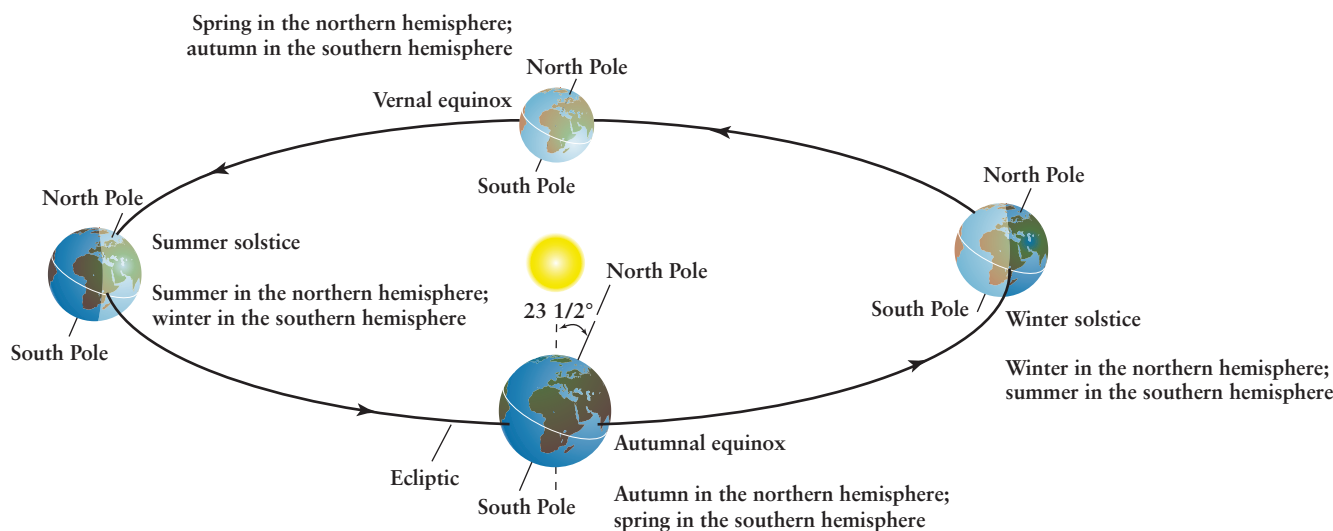


FIGURE 1-14 The Tilt of Earth’s Axis Earth’s axis of rotation is tilted $23\frac{1}{2}^\circ$ from being perpendicular to the plane of Earth’s orbit. Earth maintains this orientation (with its North Pole aimed at the north celestial pole near the star Polaris) throughout the

year as it orbits the Sun. Consequently, the amount of solar illumination and the number of daylight hours at any location on Earth vary in a regular fashion with the seasons.

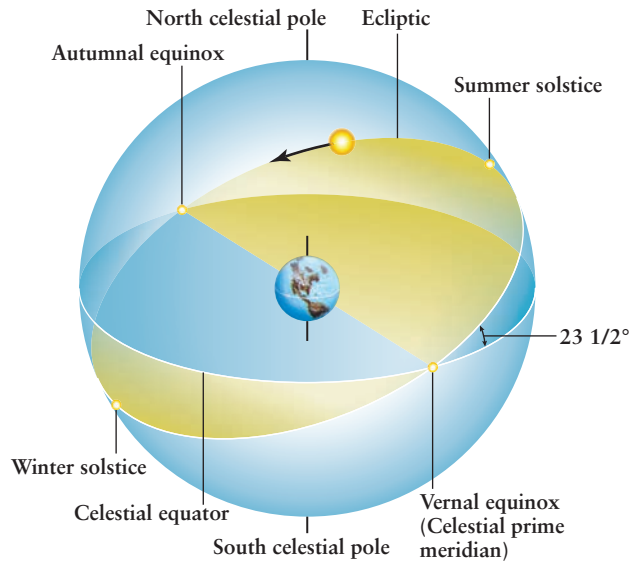


FIGURE 1-15 The Seasons Are Linked to Equinoxes and Solstices The ecliptic is inclined to the celestial equator by $23\frac{1}{2}^\circ$ because of the tilt of Earth's axis of rotation. The ecliptic and the celestial equator intersect at two points called the equinoxes. The northernmost point on the ecliptic is called the summer solstice; the southernmost point is called the winter solstice.

each year (see Figure 1-16a) and is called the **winter solstice**. The winter solstice is the point on the ecliptic farthest south of the celestial equator (see Figure 1-15). It is also the day when the Sun rises to the lowest height at noon (Figure 1-16a), and it signals the day of the year in the northern hemisphere with the fewest number of daylight hours.

As the Sun moves along the ecliptic after the winter solstice, it rises earlier, is more northerly on the eastern horizon, and it passes higher in the sky at midday than it did on preceding days. Three months later, around March 20, the Sun crosses the celestial equator heading northward. As noted in Section 1-5, this is called the *vernal equinox* and is one of the two days on which the Sun rises due east and sets due west (Figure 1-16b). The vernal equinox is the “prime meridian” of the celestial sphere. Three months after the vernal equinox, around June 21, the Sun rises farthest north of east and passes highest in the sky (Figure 1-16c). This is the **summer solstice** (see Figure 1-15), the day of the year in the northern hemisphere with the most daylight.

Each day from June 21 through December 21, the Sun rises farther south than it did the preceding day. Its highest point in the sky is lower each succeeding day—the cycle of the previous 6 months reverses. The *autumnal equinox* occurs around September 22 (Figure 1-16d), with the Sun heading southward across the celestial equator, as seen from Earth.

3 The seasons The higher the Sun rises during the day, the more daylight hours there are. During the days with longer periods of daylight, more light and heat from the Sun strike that hemisphere. Furthermore, when the Sun is higher in the sky, its energy is more concentrated on Earth's surface (see the footprint that the “cylinder of light” from the Sun makes in Figure 1-16a–d). Thus, during these days more energy is deposited on each square meter of the surface, thereby warming the surface more than when the Sun is lower in the sky. *The temperature and, hence, the seasons are determined by the duration of daylight at any place and the height of the Sun in the sky there.* (Bear in mind that winds and clouds greatly affect the weather throughout the year—we ignore these effects here.)

To summarize, the Sun is lowest in the northern sky on the winter solstice. This marks the beginning of winter in the northern hemisphere. As the Sun moves northward, the amount of daylight and heat deposited increases daily. The vernal equinox marks a midpoint in the amount of light and heat from the Sun onto the northern hemisphere and is the beginning of spring. When the Sun reaches the summer solstice, it is highest in the northern sky and is above the horizon for the most hours of any day of the year. This is the beginning of summer. Returning southward, the Sun crosses the celestial equator once again on the autumnal equinox, the beginning of fall.

4 Earth is closest to the Sun on or around January 3 of each year—the dead of winter in the northern hemisphere! While the distance between Earth and the Sun changes by 5 million km (3 million mi) throughout the year, this 3% variation in distance, and hence energy deposited from the Sun, is not great enough to cause the seasons; it has only a minor effect on them.

The Sun's path across the sky During the northern hemisphere's summer months, when the northern hemisphere is tilted toward the Sun (see Figure 1-14), the Sun rises in the northeast and sets in the northwest. The Sun provides more than 12 hours of daylight in the northern hemisphere and passes high in the sky. At the summer solstice, the Sun is as far north as it gets, giving the greatest number of daylight hours to the northern hemisphere (Figure 1-16c).

Focus Question 1-3

Explain why Figure 1-12 must have been taken facing west.

Hint: Examine Figure 1-16.

Insight Into Science

Expect the Unexpected The process of science requires that we question the obvious, that is, what we think we know. Many phenomena in the universe defy commonsense explanations. The fact that the changing distance from Earth to the Sun has a minimal effect on the seasons is an excellent example.

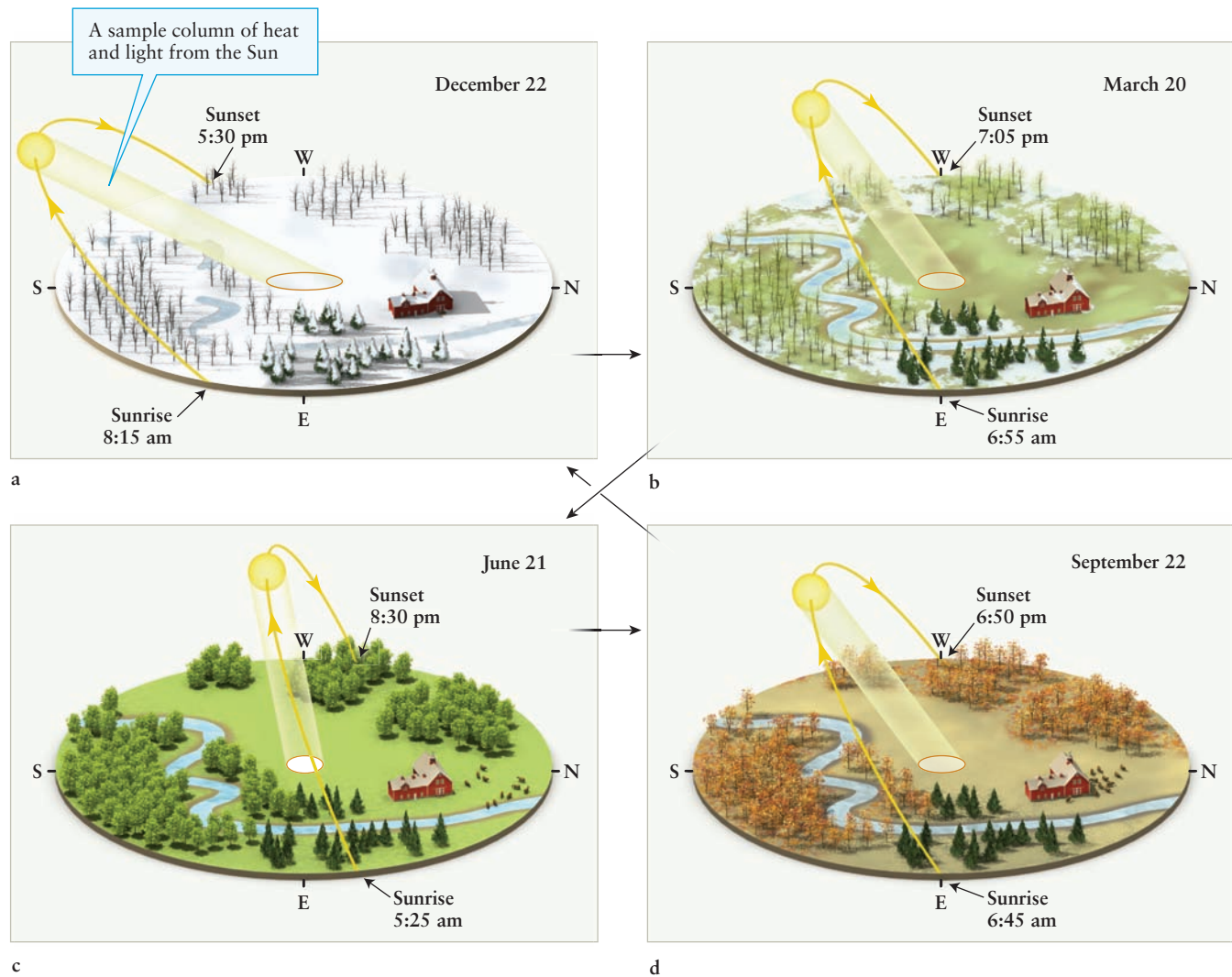


FIGURE 1-16 The Sun's Daily Path and the Energy It Deposits Here Northern Hemisphere: (a) On the winter solstice—the first day of winter—the Sun rises farthest south of east, is lowest in the noontime sky, stays up the shortest time, and its light and heat are least intense (most spread out) of any day of the year in the northern hemisphere. (b) On the vernal equinox—the first day of spring—the Sun rises precisely in the east and sets precisely in the west. Its light and heat have been growing more intense, as shown by the brighter oval of light than in (a). (c) On the summer solstice—the first day of summer—the Sun rises farthest north of east of any day in the year, is highest in the noontime sky, stays up the longest

time, and its light and heat are most intense of any day in the northern hemisphere. (d) On the autumnal equinox, the same astronomical conditions exist as on the vernal equinox. **South-ern Hemisphere:** If you are reading this in the southern hemisphere, make the following changes: (a) Change December 22 to June 21 and visualize the Sun's path starting and ending the same distance *north* of east and north of west as it is south of east and south of west as shown here; (b) change March 20 to September 23; (c) change June 21 to December 21 and visualize the Sun's path starting and ending the same distance *south* of east and south of west as shown north of east and north of west here; (d) change September 22 to March 20.

During the northern hemisphere's winter months, when the northern hemisphere is tilted away from the Sun, the Sun rises in the southeast. Daylight lasts for fewer than 12 hours, as the Sun skims low over the southern horizon and sets in the southwest. Night is longest in the northern hemisphere when the Sun is at the winter solstice (Figure 1-16a).

The Sun's maximum angle above the southern horizon is different at different latitudes. The farther north

you are, the lower the Sun is in the sky at any time of day than it is on that day at more equatorial locations. At latitudes above $66\frac{1}{2}^\circ$ north latitude or below $66\frac{1}{2}^\circ$ south latitude, the Sun does not rise above the horizon at all during parts of their fall and winter months. During their spring and summer months, those same regions of Earth have continuous sunlight for weeks or months (Figure 1-17) because the Sun is then circumpolar; hence, the name "Land of the Midnight Sun."

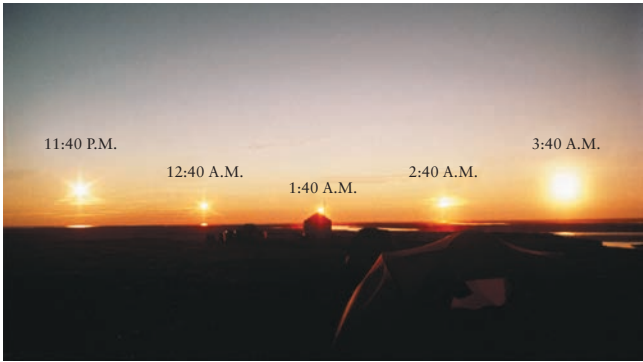


FIGURE 1-17 The Midnight Sun This time-lapse photograph was taken on July 19, 1985, at 69° north latitude in northeastern Alaska. At that latitude, the Sun is above the horizon continuously from mid-May until the end of July. (Doug Plummer/ Science Source)

5 The Sun takes 1 year to complete a trip around the ecliptic (as noted above, this motion is actually caused by Earth’s orbit around the Sun). Since there are about $365\frac{1}{4}$ days in a year and 360° in a circle, the Sun appears to move along the ecliptic at a rate of slightly less than 1° per day. The constellations through which the Sun moves throughout the year as it travels along the ecliptic are called the **zodiac** constellations. We cannot see the stars of these constellations when the Sun is among them (see Figure 1-13), but we can plot the Sun’s path on the celestial sphere to determine through which constellations it moves. Traditionally, there were 12 zodiac constellations whose borders were set in antiquity. Recall from Section 1-3 that in 1930 the boundaries of the constellations were redefined by astronomers. As a result the Sun now moves through 13 constellations throughout the year. (The thirteenth zodiac constellation is Ophiuchus, the Serpent Holder. The Sun passes through Ophiuchus from December 1 to December 19 each year.) **Table 1-1** lists all the zodiac constellations and the dates the Sun passes through them. You may not have the “sign” that you think you have.

1-8 Clocks and calendars are based on Earth’s rotation and revolution

The Sun’s daily motion through the sky provided our distant ancestors’ earliest reference for time because the Sun’s location determines whether it is day or night and roughly whether it is before or after midday. The Sun’s motion through the sky came to determine the length of the **solar day**, upon which our 24-hour day is based. Ideally, this is the interval of time between when the Sun is highest in the sky on one day until the time it is highest in the sky on the next day. However, the length of the solar day actually varies throughout the year, up to 30 seconds per day. This occurs because Earth’s orbit around the Sun is not perfectly circular—our planet’s

TABLE 1-1 THE 13 CONSTELLATIONS OF THE ZODIAC

Constellation	Dates of the Sun’s passage through each
Pisces	March 13–April 20
Aries	April 20–May 13
Taurus	May 13–June 21
Gemini	June 21–July 20
Cancer	July 20–August 11
Leo	August 11–September 18
Virgo	September 18–November 1
Libra	November 1–November 22
Scorpius	November 22–December 1
Ophiuchus	December 1–December 19
Sagittarius	December 19–January 19
Capricorn	January 19–February 18
Aquarius	February 18–March 13

speed along the ecliptic increases as we approach the Sun and it decreases as we move away from it—and because Earth’s rotation axis is tilted $23\frac{1}{2}^\circ$ from being perpendicular to the ecliptic. These two effects change the apparent speed of the Sun across the sky from day to day. The *average* time interval between consecutive noontimes throughout the year is 24 hours, which determines the time we use on our clocks. This is called the *mean* (or *average*) solar day.

Along with the solar day, there is also a *sidereal day*, the length of time from when any star is in one place in the sky until it is next in the same place. The solar and sidereal days differ from each other in length because while Earth rotates, it also revolves around the Sun. This motion of Earth in its orbit day by day changes the location of the stars, bringing them back to their original positions 4 minutes earlier each day, as noted in Section 1-6. Therefore, the sidereal day is 23 hours, 56 minutes, 4 seconds long, while the solar day is 24 hours long (**Figure 1-18**).

Just as the day is caused by Earth’s rotation, the year is the unit of time based on Earth’s revolution about the Sun. As noted earlier, Earth does not take exactly 365 days to orbit the Sun, so the year is not exactly 365 days long. Basing the year on a 365-day cycle led to important events (such as holidays) occurring on the wrong day. To resolve this problem, Roman statesman Julius Caesar implemented a new calendar in the year 46 B.C.E. Because measurement revealed to ancient astronomers that the length of a year is approximately $365\frac{1}{4}$ days, this “Julian” calendar established a system of leap years

Focus Question 1-4
What region of Earth has the smallest range of seasonal temperature changes and why?

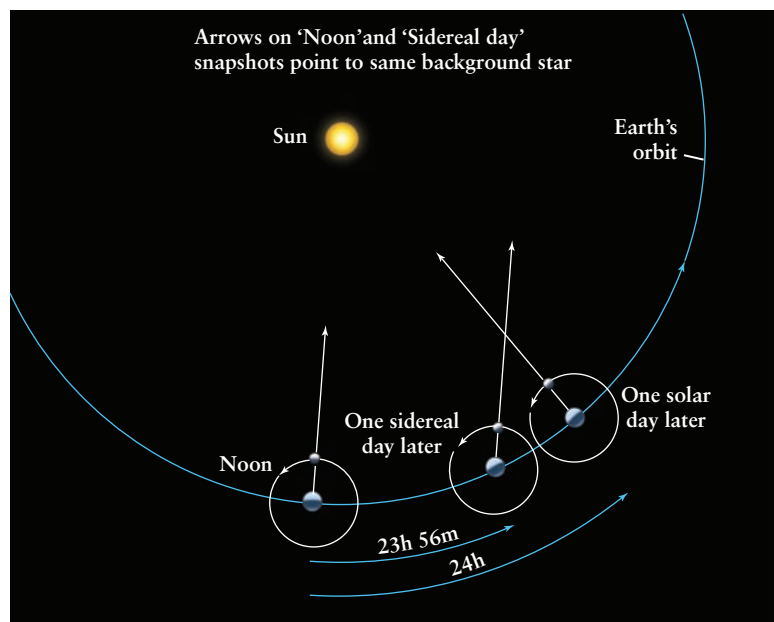


FIGURE 1-18 Sidereal and Solar Days Combining Earth's rotation and its orbit around the Sun, the length of the sidereal day can be seen (on average) to be 4 minutes shorter than the solar day. The daily motion of Earth is exaggerated here for clarity.

to accommodate the extra quarter of a day. By adding an extra day to the calendar every 4 years, Caesar hoped to ensure that seasonal astronomical events, such as the beginning of spring, would occur on the same date year after year.

The Julian calendar would have worked just fine if a year were exactly $365\frac{1}{4}$ days long and if Earth's rotation axis (currently pointing toward Polaris, as discussed earlier) never changed direction. However,

neither assumption is correct. Thus, over time, a discrepancy accumulated between the calendar and the actual time—astronomical and cultural events continued to fall on different dates each year. To straighten things out, a committee established by Pope

Gregory XIII recommended a refinement, thus creating the Gregorian calendar in 1582. In the Gregorian system, which we use today, we have a leap year every four years with the exception that century years are leap years *only* if evenly divisible by 400. For example, the years 1700, 1800, and 1900 were not leap years under the improved Gregorian system. But the year 2000—which can be divided evenly by 400—was a leap year. The Gregorian system assumes that the year is 365.2425 mean solar days long, which is very close to the length of the *tropical year*, defined as the time interval from one vernal equinox to the next. In fact, the error is only 1 day in every 3300 years. That will not cause any problems for a long time.

Focus Question 1-5

Why would a calendar based on sidereal days not be satisfactory?

1-9 Precession is a slow, circular motion of Earth's axis of rotation

As just noted above, the orientation of Earth's axis of rotation changes slightly with respect to the celestial sphere (that is, it “points” in a slightly different direction over time). While this change is small over our lifetimes, it eventually causes the north celestial pole to drift away from Polaris. This major change in orientation is caused by gravitational forces from the Moon and the Sun pulling on the slight bulge at Earth's equator created by Earth's rotation—our planet's diameter is about 43 km (27 mi) greater at the equator than from pole to pole. **Gravitation (gravity)** is the universal force of attraction between all matter. The strength of the gravitational force between two objects depends on the amounts of mass the objects have and the distance between them, as we explain in more detail in Chapter 2.

Because of Earth's tilted axis of rotation, the Sun and Moon are usually not located directly over Earth's equator. As a result, their gravitational attractions on Earth's equatorial bulge provide forces that pull the bulge toward them (**Figure 1-19a**). However, Earth does not respond to these forces from the Sun and Moon by tilting so that its equator is closest to them. Instead, Earth maintains the same tilt ($23\frac{1}{2}^\circ$ from the ecliptic), but the direction in which its axis of rotation points on the celestial sphere changes—a motion called **precession**. This is exactly the same behavior exhibited by a spinning top (**Figure 1-19b**). If the top were not spinning, gravity would pull it over on its side. But when it is spinning, the combined actions of gravity and rotation cause the top's axis of rotation to precess or wobble in a circular path. As with the toy top, the combined actions of gravity from the Sun and the Moon plus rotation cause Earth's axis to trace a circle in the sky while remaining tilted about $23\frac{1}{2}^\circ$ away from the ecliptic.

In the mid-1990s, astronomers simulated the behavior of Earth and discovered that without a large moon, Earth would not keep to a $23\frac{1}{2}^\circ$ tilt, but rather would change its angle relative to the ecliptic dramatically over millions of years. Therefore, while being a major cause of precession, the Moon plays a crucial role in stabilizing Earth and maintaining the seasons as we know them.

Earth's rate of precession is slow compared to human timescales. It takes about 26,000 years for the north celestial pole to trace out a complete circle around the sky, as shown in **Figure 1-19c**. (The south celestial pole executes a similar circle in the southern sky.) At the present time, Earth's north axis of rotation points within 1° of the star Polaris. In 3000 B.C.E., it pointed near the star Thuban in the constellation Draco (the Dragon). In the year 14,000, the pole star will be near Vega in Lyra. As Earth's axis of rotation precesses, its equatorial plane also moves. Because Earth's equatorial plane defines the location of the celestial equator in the sky, the celestial

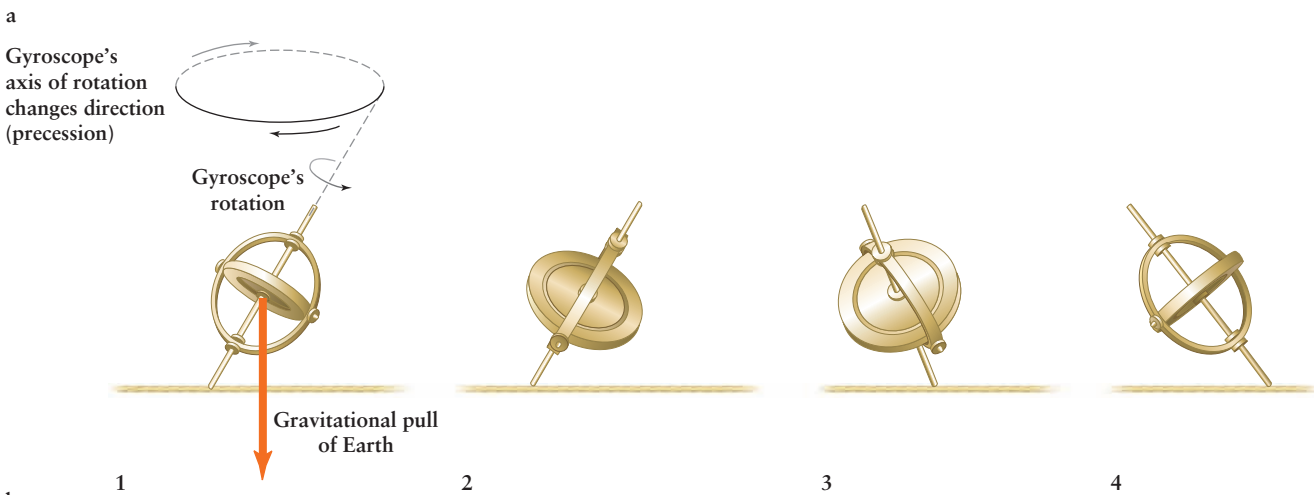
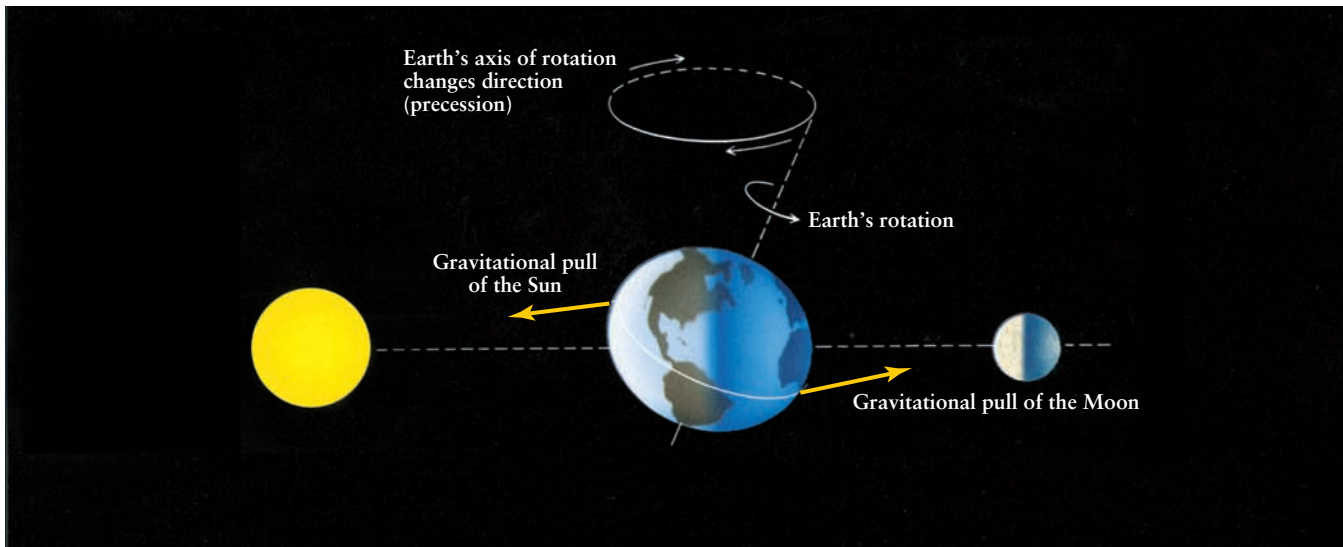


FIGURE 1-19 Precession and the Path of the North Celestial Pole (a) The gravitational pulls of the Moon and the Sun on Earth's equatorial bulge cause Earth to precess. (b) The situation is analogous to the motion of a gyroscope. The top of the gyroscope shows the motion of Earth's North Pole or South Pole, while the point on which the gyroscope spins represents the center of Earth. As the gyroscope spins, Earth's gravitational pull causes the gyroscope's axis of rotation to move in a circle—to precess. (c) As Earth precesses, the north celestial pole slowly traces out a circle among the northern constellations. At the present time, the north celestial pole is near the moderately bright star Polaris, which serves as the pole star. The total precession period is about 26,000 years.

equator also changes over time. Recall that the intersections of the celestial equator and the ecliptic define the equinoxes, so these key locations in the sky shift slowly from year to year. This entire phenomenon is often called the **precession of the equinoxes**. This change was discovered by the great Greek astronomer Hipparchus. Today, the vernal equinox is located in the constellation Pisces (the Fish). Two thousand years ago, it was located in Aries (the Ram). Around the year 2600, the vernal equinox will move into Aquarius (the Water Bearer).

THE PHASES OF THE MOON

1-10 The phases of the Moon originally inspired the concept of the month

While the Moon's precessional effect on Earth unfolds over millennia, many lunar effects are noticeable every day. As the Moon orbits Earth, it moves from west to east (right to left) on the celestial sphere, changing position among the background stars. Its position relative to the Sun also changes and, as a result, we see different **lunar phases**.

The Sun illuminates half of the Moon at all times (**Figure 1-20**). The Moon's phase that we see depends on how much of its sunlit hemisphere is facing Earth. When the Moon is closest to the Sun in the sky, its dark hemisphere faces Earth. This phase, during which the Moon is at most a tiny crescent, is called the *new* Moon. During the 7 days following the new phase, more of the Moon's illuminated hemisphere becomes exposed to our view, resulting in a phase called the *waxing crescent* Moon. At the *first quarter* Moon, we see half of the illuminated hemisphere and half of the dark hemisphere. "Quarter Moon" refers to how far in its cycle the Moon has gone, rather than what fraction of the Moon appears lit by sunlight.

During the next week, still more of the illuminated hemisphere can be seen from Earth, giving us the phase called the *waxing gibbous* Moon. "Gibbous" means "rounded on both sides." When the Moon arrives on the opposite side of Earth from the Sun, we see virtually all the fully illuminated hemisphere. This phase is the *full* Moon. Over the following 2 weeks, we see less and less of the illuminated hemisphere as the Moon continues along its orbit. This movement produces the phases called the *waning gibbous* Moon, the *third quarter* Moon, and the

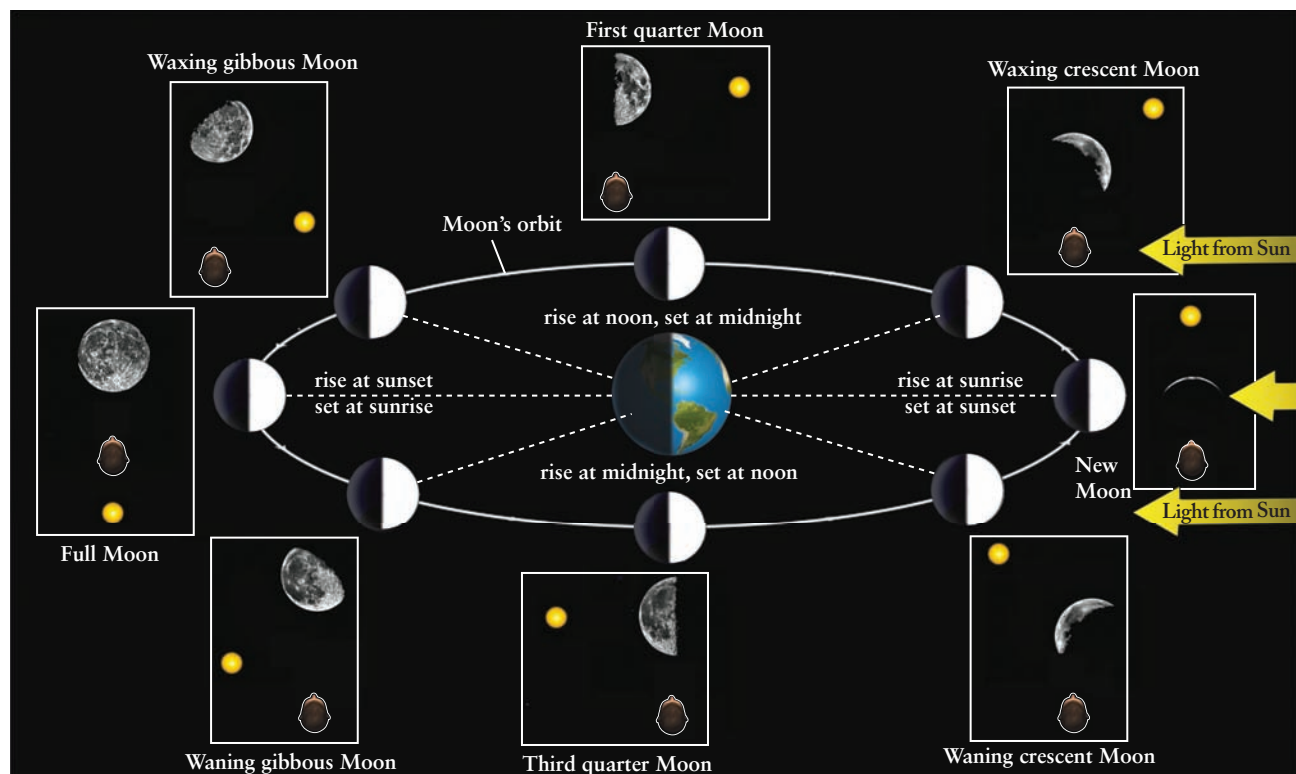


FIGURE 1-20 The Phases of the Moon The diagram shows the Moon at eight locations on its orbit as viewed from far above Earth's North Pole. Light from the Sun illuminates one-half of the Moon at all times, while the other half is dark. It

takes about $29\frac{1}{2}$ days for the Moon to go through all its phases. The inset drawings with photographs show the resulting lunar phases as seen from Earth. (Yerkes Observatory and Lick Observatory)

waning crescent Moon. The Moon completes a full cycle of phases in $29\frac{1}{2}$ days.

6 Confusion often occurs over the terms “far side” and “dark side” of the Moon. The far side is the side of the Moon facing away from Earth. The dark side is the side of the Moon on which the Sun is not shining. By examining the photographs in Figure 1-20 you can see that the same physical side of the Moon faces Earth all the time (that is, we always see the same craters). The half of the Moon that never faces Earth is the far side. However, the far side is not always the dark side, because we see part of the dark side whenever we see less than a full Moon. (What did Pink Floyd have to say about the dark side of the Moon?)

Figure 1-20 shows the Moon at various positions in its orbit. Remember that the bright side of the Moon is on the right (west) side of the waxing Moon, while the bright side is on the left (east) side of the waning Moon. This information can tell you at a glance whether the Moon is waxing or waning. When looking at the Moon through a telescope, the best place to see surface details is where the



R I V U X G

FIGURE 1-21 The Moon During the Day The Moon is visible at some time during daylight hours virtually every day. The time of day or night it is up in our sky depends on its phase. (Richard Cummins/SuperStock)

shadows are longest. This occurs at the boundary between the bright and dark regions, called the **terminator**.

7 Figure 1-20 also shows local time around the globe when the Moon rises and sets at the indicated phases. For example, at first quarter, the Moon is 90° east of the Sun on the celestial sphere; hence, the Moon rises when the Sun is highest in the sky. At full Moon, the Moon is opposite the Sun in the sky; thus, the Moon rises as the Sun is setting. Using this information, you can see that the Moon is visible during the daytime (**Figure 1-21**) for a part of most days of the year.

The approximately 4 weeks that the Moon takes to complete one cycle of its phases inspired our ancestors to invent the concept of a month. Astronomers find it useful to define two types of months, depending on whether the Moon’s motion is measured relative to the stars or to the Sun. Neither type corresponds exactly to the months of our usual calendar, which have different (and, in the case of February, varying) lengths.

The **sidereal month** is the time it takes the Moon to complete one full orbit of 360° around Earth (**Figure 1-22**). The length of the sidereal month is determined

Focus Question 1-6

Is the Moon waning or waxing in Figure 1-21?

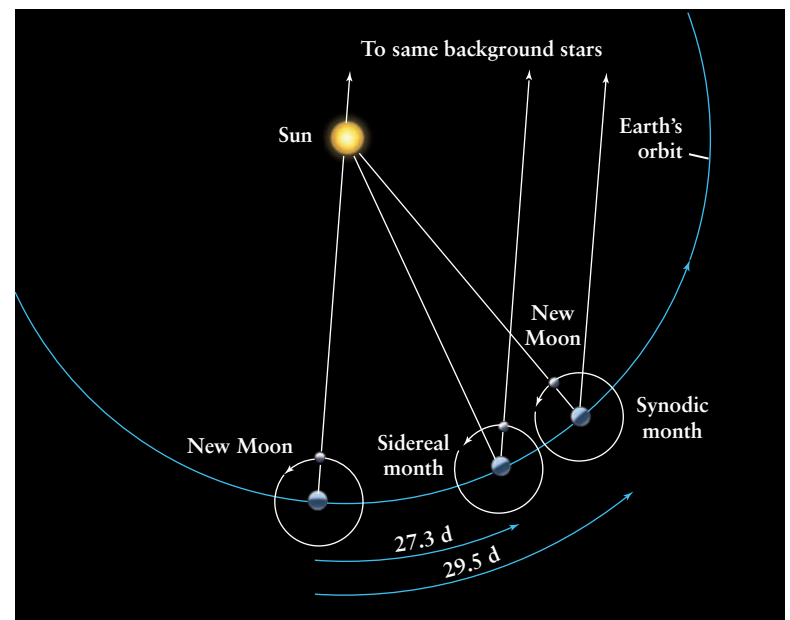


FIGURE 1-22 The Sidereal and Synodic Months The sidereal month is the time it takes the Moon to complete one revolution with respect to the background stars, about 27.3 days. However, because Earth is constantly moving in its orbit about the Sun, the Moon must travel through more than 360° to get from one new Moon to the next. The synodic month is the time between consecutive new Moons or consecutive full Moons, about $29\frac{1}{2}$ days.

by the location of the Moon in its orbit around Earth as measured with respect to the stars, analogous to how the sidereal day is measured. Equivalently, this is the time it takes the Moon to start at one place on the celestial sphere and return to exactly the same place again. The sidereal orbital period of the Moon takes approximately 27.3 days. The **synodic month**, or **lunar month**, is the time it takes the Moon to complete one 29½-day cycle of phases (for example, from new Moon to new Moon or from full Moon to full Moon) and thus is measured with respect to the Sun rather than the stars.

The synodic month is longer than the sidereal month because Earth is orbiting the Sun while the Moon goes through its phases. As shown in Figure 1-22, the Moon must travel *more* than 360° along its orbit around Earth to complete a cycle of phases (for example, from one new Moon to the next), which takes about 2.2 days longer than the sidereal month.

Both the sidereal month and synodic month vary somewhat, because the gravitational pull of the Sun on Earth changes Earth's orbital speed throughout the year and the gravitational pull of the Sun on the Moon affects the Moon's speed as it orbits Earth. The sidereal month can vary by as much as 7 hours, while the synodic month can change by as much as 12 hours.

The terms *synodic* and *sidereal* are also used in discussing the motion of the other bodies in the solar system. The synodic period of a planet is the time between consecutive straight alignments of the Sun, Earth, and that planet (during which time interval the planet also goes through a cycle of phases, as seen from Earth). Recall that any orbit measured with respect to the stars is called “sidereal,” including orbits of the planets around the Sun, as well as orbits of moons around their planets.

ECLIPSES

8 Eclipses are among the most spectacular natural phenomena. During a **lunar eclipse**, the brilliant full Moon often darkens to a rusty red. A lunar eclipse occurs when the Moon passes through Earth's shadow. This can happen only when the Sun, Earth, and the Moon are in a straight line at full Moon. During a **solar eclipse**, broad daylight is transformed into an eerie twilight, as the Sun seems to be blotted from the sky. A solar eclipse occurs when the Moon's shadow moves across Earth's surface. As seen from Earth, the Moon moves in front of the Sun—a new Moon.

1-11 Eclipses do not occur during every new or full Moon phase

At first glance, it would seem that eclipses should happen at every new and full Moon, but in fact, they occur much less often because the Moon's orbit is tilted 5° from the plane of the ecliptic (**Figure 1-23**). Consequently, the new Moon and full Moon usually occur when the Moon is either above or below the plane of Earth's orbit. In such positions, a true alignment between the Sun, Moon, and Earth is not possible. Hence, Earth and the Moon do not enter each others' shadows and an eclipse cannot occur. This tilt also explains why, as mentioned earlier, the new Moon is often seen as a thin crescent—we see “under” or “above” it to a tiny sliver of the sunlit side.

Indeed, because its orbit is tilted 5° from the ecliptic, the Moon is usually above or below the plane of our orbit around the Sun. The Moon crosses the ecliptic at what is called the **line of nodes** (see Figure 1-23). An eclipse takes place only when the Moon crosses the plane of the ecliptic during its new or full phase. By cal-

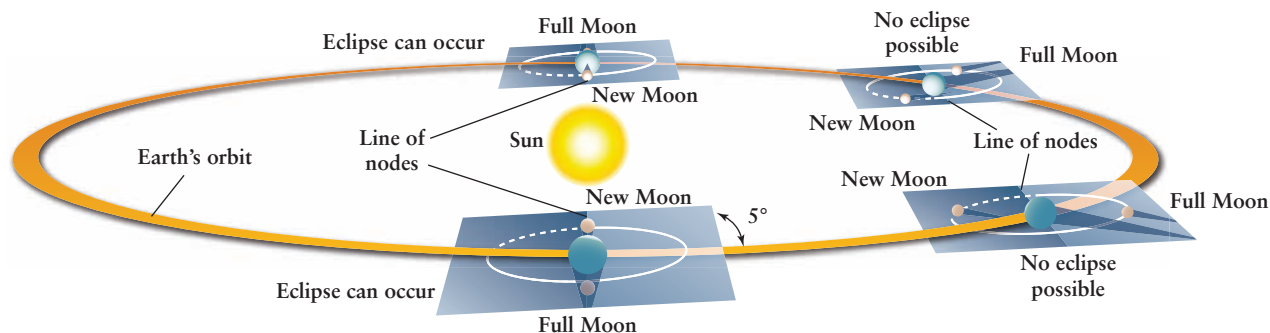
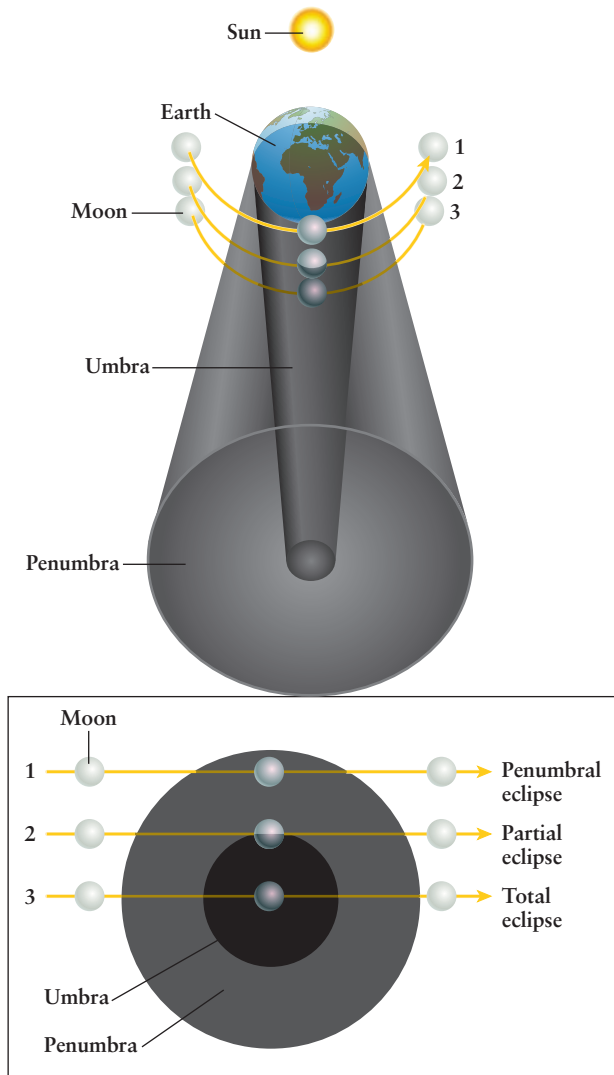
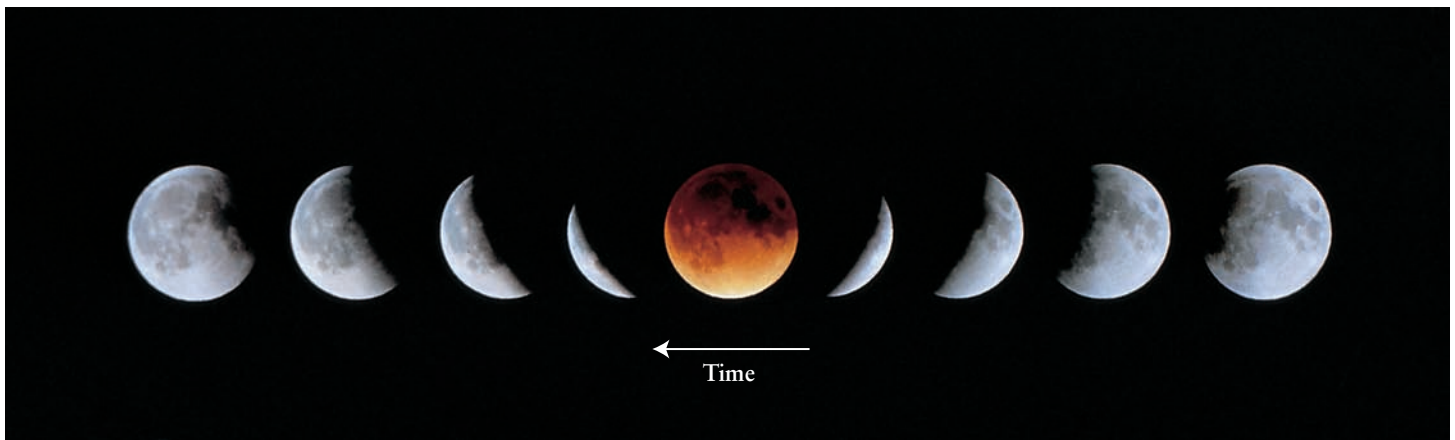


FIGURE 1-23 Conditions for Eclipses The Moon must be very nearly on the ecliptic at new Moon for a solar eclipse to occur. A lunar eclipse occurs only if the Moon is very nearly on

the ecliptic at full Moon. When new Moon or full Moon phases occur away from the ecliptic, no eclipse is seen because the Moon and Earth do not pass through each other's shadows.



a



b

RITVUXIG

FIGURE 1-24 Three Types of Lunar Eclipses (a) People on the nighttime side of Earth see a lunar eclipse when the Moon moves through Earth's shadow. The umbra is the darkest part of the shadow. In the penumbra, only part of the Sun is covered by Earth. The inset shows the various lunar eclipses that occur,

culating the number of times a new Moon takes place on the line of nodes, we find that at least two and no more than five solar eclipses occur each year. Lunar eclipses occur just about as frequently as solar eclipses, with the maximum number of eclipses (solar plus lunar) possible in a year being seven.

1-12 Three types of lunar eclipses occur

Earth's shadow has two distinct parts, as shown in **Figure 1-24a**. The **umbra** is the part of the shadow where all direct sunlight is blocked by Earth. If you were in Earth's umbra looking at Earth, you would not see the Sun behind it at all. The **penumbra** of the shadow is where Earth blocks only some of the sunlight. If you were in Earth's penumbra looking at Earth, you would see a crescent Sun behind it. The Moon has an analogous umbra and penumbra.

Depending on how the Moon travels through Earth's shadow, one of three kinds of lunar eclipse may occur. A **penumbral eclipse**, when the Moon passes through only Earth's penumbra, is easy to miss. The Moon still looks full, just a little dimmer than usual and sometimes slightly reddish in color (path 1 in Figure 1-24a).

When just part of the lunar surface passes through the umbra, a bite seems to be taken out of the Moon, and we see a **partial eclipse** (path 2 in Figure 1-24a). When the Moon travels completely into the umbra, we see a **total eclipse** of the Moon (path 3 in Figure 1-24a). Total lunar eclipses with the maximum duration, lasting for up to 1 hour and 47 minutes, occur when the Moon is closest to Earth and is traveling directly through the center of the umbra.

depending on the Moon's path through Earth's shadow. (b) This sequence of nine photographs was taken over a 3-hour period during the total lunar eclipse of January 20, 2000. During the total phase, the Moon has a distinctly reddish color. (© Fred Espenak, MrEclipse.com)

TABLE 1-2 LUNAR ECLIPSES, 2015–2018

Date	Visible from	Type	Duration of totality (h:min)
2015 Apr 04	Asia, Australia, Pacific, Americas	Total	:05
2015 Sep 28	Europe, Africa, west Asia, Americas	Total	1:12
2016 Mar 23	Asia, Australia, Pacific, western Americas	Penumbral	
2016 Sep 16	Europe, Africa, Asia, Australia	Penumbral	
2017 Feb 11	Europe, Africa, Asia, Americas	Penumbral	
2017 Aug 07	Europe, Africa, Asia, Australia	Partial	
2018 Jan 31	Asia, Australia, western North America	Total	1:16
2018 Jul 27	South America, Europe, Africa, Asia, Aus.	Total	1:43

Table 1-2 lists all the total and partial lunar eclipses from 2015 through 2018.

Even during a total eclipse, the Moon does not completely disappear. A small amount of sunlight passing through Earth’s atmosphere is bent into Earth’s umbra.

The light deflected into the umbra is primarily red and orange, and thus the darkened Moon glows faintly in rust-colored hues (Figure 1-24b). Everyone on the side of Earth over which a lunar eclipse occurs can see it, provided that clouds do not obscure the event.

Lunar eclipses are perfectly safe to watch with the naked eye.

Focus Question 1-7
Why does the new Moon sometimes appear as a crescent?

1-13 Three types of solar eclipses also occur

The different physical diameters of the Sun and Moon and their different distances from us combine so that they both have nearly the same angular diameter as seen from Earth—about $\frac{1}{2}^\circ$. When the Moon completely covers the Sun, the result is a total solar eclipse. You can see in Figure 1-25 that only the tip of the Moon’s umbra ever reaches Earth’s surface. As Earth turns and the Moon orbits, the tip traces an **eclipse path** across our planet. Only people within these areas are treated to the spectacle of a total solar eclipse, which is why cruise lines love solar eclipses. During those few precious moments, hot gases (the **solar corona**) surrounding the Sun can be observed and photographed (Figure 1-26). By studying the light from the outer layers of the Sun, astronomers have been able to learn about its atmospheric temperature, chemistry, and gas activity.

Viewing the Sun directly for more than a moment at any time without an approved filter causes permanent eye damage. It is only safe to look at a total solar eclipse

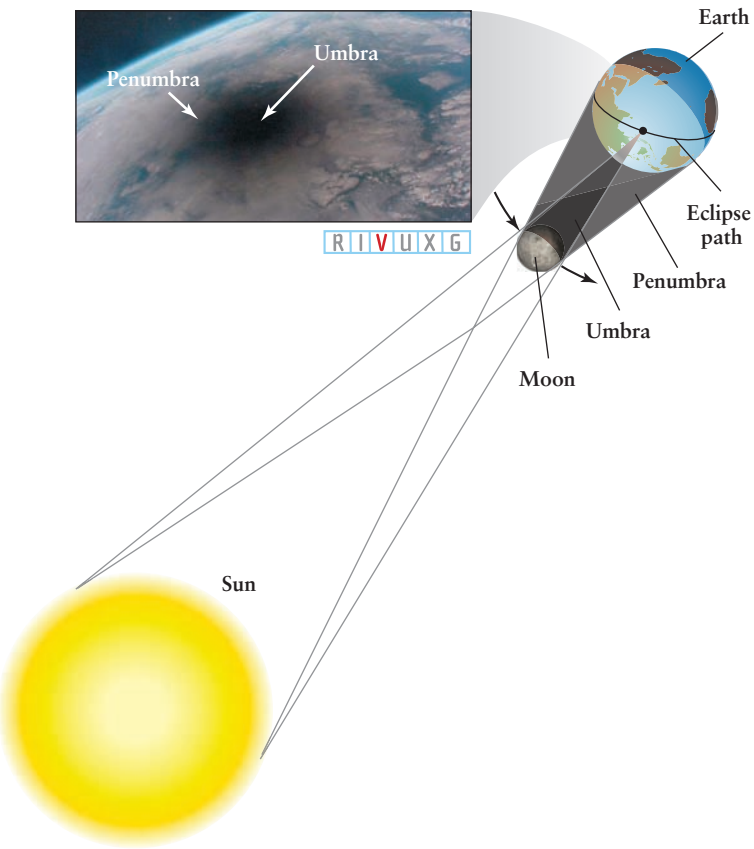
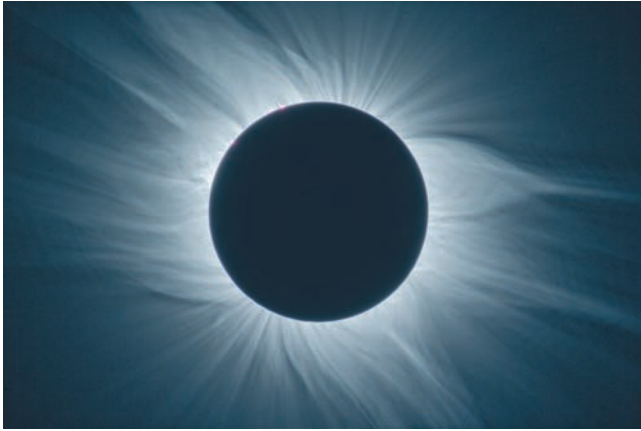


FIGURE 1-25 The Geometry of a Total Solar Eclipse During a total solar eclipse, the tip of the Moon’s umbra traces an eclipse path across Earth’s surface. People inside the eclipse path see a total solar eclipse, while people inside the penumbra see only a partial eclipse. The photograph in this figure shows the Moon’s shadow on Earth. It was taken from the *Mir* space station during the August 11, 1999, total solar eclipse. The Moon’s umbra appears as the very dark spot on the eastern coast of the United States. The umbra is surrounded by the penumbra. (Inset: NASA/SSPL/Getty Images)



RIVUXG

FIGURE 1-26 A Total Eclipse of the Sun During a total solar eclipse, the Moon completely covers the Sun's disk, and the solar corona can be photographed. This halo of hot gases extends for millions of kilometers into space. This gorgeous image was taken in southwestern Mongolia during the August 1, 2008, solar eclipse. (Atlas Photo Bank/Science Source)

without a filter during the brief time when the Moon completely blocks the Sun. At all other times during the eclipse, you must view it either through an approved filter or in the image made on a flat surface by a telescope or a pinhole camera.

Earth's rotation and the orbital motion of the Moon cause the umbra to race along the eclipse path at speeds in excess of 1700 km/h (1050 mi/h). For this reason, complete blockage of the Sun (totality) during a total eclipse never lasts for more than 7½ minutes at any one location on the eclipse path, and it usually lasts for only a few moments.

The Moon's umbra is also surrounded by a penumbra (see Figure 1-25). The photograph in Figure 1-25 shows the dark spot produced by the Moon's umbra and the less dark penumbral ring surrounding it on Earth's surface during a total solar eclipse. During a solar eclipse, the Moon's penumbra extends over a large portion of Earth's surface. When *only* the penumbra sweeps across Earth's surface, as happens in high latitude regions, the Sun is only covered by the Moon. This circumstance results in a *partial eclipse of the Sun*. Similarly, people in the penumbra of a total eclipse see a partial eclipse. In either case, the Sun looks crescent-shaped as seen from Earth. *Never look directly at a partial solar eclipse.*

The Moon's orbit around Earth is oval shaped. Therefore, the distance between Earth and the Moon, which averages 384,400 km (238,900 mi), varies by 14% as the Moon moves around Earth. Consequently, the angular size of the Moon in our sky varies by that amount. The width of the eclipse path depends primarily on the Earth–Moon distance during an eclipse. The eclipse path is widest—up to 270 km (170 mi)—when the new Moon happens to be at the point in its orbit

nearest Earth. Usually, however, the path is much narrower.

If what would otherwise be a total solar eclipse occurs when the Moon is relatively far from Earth, then the Moon's umbra falls short of Earth and no one sees a total eclipse. From Earth's surface, the Moon then appears too small to cover the Sun completely, and a thin ring or “annulus” of light is seen around the edge of the Moon at mid-eclipse. This type of eclipse is called an **annular eclipse** (Figure 1-27). The length of the Moon's umbra is nearly 5000 km (3100 mi) shorter than the average distance between the Moon and Earth's surface. Thus, the Moon's shadow often fails to reach Earth, making annular solar eclipses slightly more common than total solar eclipses. Table 1-3 lists all the total, partial, and annular solar eclipses from 2015 through 2018. Figure 1-28 shows the paths over Earth of all the total and annular eclipses through 2020. Please remember: It is *never* safe to look directly at a partial or annular eclipse.

Focus Question 1-8

Why aren't there any annular lunar eclipses?



RIVUXG

FIGURE 1-27 An Annular Eclipse of the Sun This composite of five exposures taken at sunrise in Costa Rica shows the progress of an annular eclipse of the Sun that occurred on December 24, 1974. Note that at mid-eclipse the edge of the Sun is visible around the Moon. (Dennis Di Cicco)

TABLE 1-3 SOLAR ECLIPSES, 2015–2018

Date	Type	Visible from	Total/annular eclipse time (min:sec)
2015 Mar 20	Partial/Total	Partial: Iceland, Europe, northern Africa, northern Asia Total: north Atlantic Ocean	2:47
2015 Sep 13	Partial	Southern Africa, Antarctica	
2016 Mar 09	Partial/Total	Partial: eastern Asia, Australia Total: Pacific Ocean, Borneo	4:09
2016 Sep 01	Annular	Atlantic Ocean, central Africa, Indian Ocean	3:06
2017 Feb 26	Annular	Pacific Ocean, Chile, Argentina, Atlantic Ocean, Africa	0:44
2017 Aug 21	Partial/Total	Partial: northern South America, all of North America except United States Total: United States	2:40
2018 Feb 15	Partial	southern South America	
2018 Jul 13	Partial	southern Australia	
2018 Aug 11	Partial	northern Europe, northeast Asia	

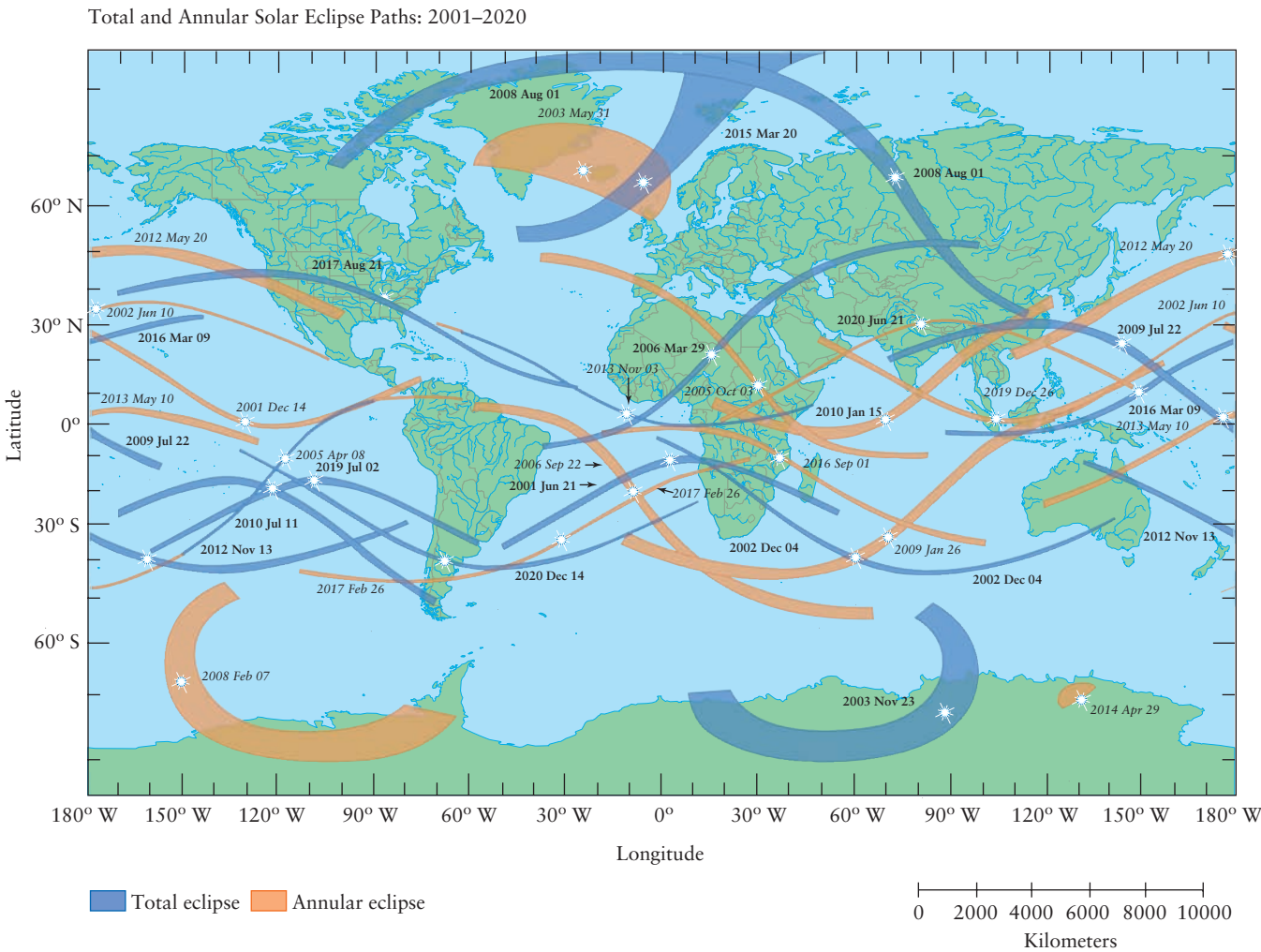


FIGURE 1-28 Eclipse Paths for Total and Annular Eclipses 2001–2020 This map shows the eclipse paths for the 14 total solar and 13 annular eclipses occurring between 2001 and

2020. In each eclipse, the Moon’s shadow travels along the eclipse path in a generally eastward direction across Earth’s surface. (Courtesy of Fred Espenak, NASA/Goddard)

SCALES OF THE UNIVERSE

Now you are oriented to what you see in the night sky, but there is so much more to astronomy, as you'll see in the remaining chapters of this book. In learning a new field, it is often useful to see the “big picture” before exploring the details. For context, here we briefly survey the major types of objects in the universe, along with their ranges of size and the scale of the distances between them.

1-14 Astronomical distances are, well, astronomical

One of the thrills and challenges of studying astronomy is becoming familiar and comfortable with the vast range of sizes that occurs in it. In our everyday lives we typically deal with distances ranging from millimeters to thousands of kilometers. (The metric or International System (SI) of units is standard in science and will be used throughout this book; however, we will often provide the equivalent in U.S. customary units. Appendix K: Common Conversions lists conversion factors between these two sets of units.)

A hundredth of a meter or a thousand kilometers are numbers that are easy to visualize and write. In astronomy, we deal with particles as small as a millionth of a billionth of a meter and systems of stars as large as a thousand billion billion kilometers across. Similarly, the speeds of some things, like light, are so large as to be cumbersome if you have to write them out in words each time. **Scientific notation** (Appendix A: Powers-of-Ten Notation) makes comparisons easy, telling us how many factors of 10 in size, mass, brightness, distance, and other parameters one object is compared to another.

The size of the observable universe and the range of sizes of the objects in it are truly staggering. **Figure 1-29** summarizes the array of sizes from atomic particles up to the diameter of the entire universe visible to us. Unlike linear intervals measured on a ruler, moving up 0.5×10^{-2} (0.5) cm along the arc of this figure brings you to objects 10 times larger. Because of this, going from the size of a proton (roughly) up to the size of an atom (roughly) takes about the same space along the arc as going from the distance between Earth and

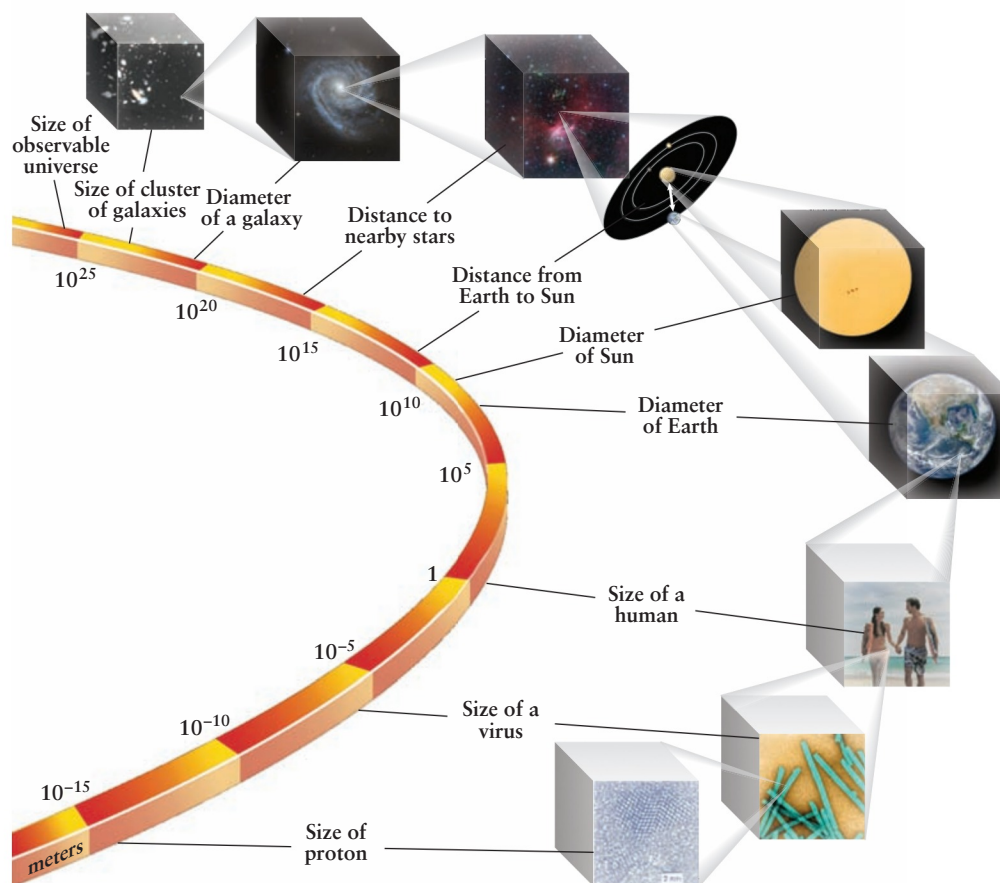


FIGURE 1-29 The Scales of the Universe This curve gives the sizes of objects in meters, ranging from subatomic particles at the bottom to the entire observable universe at the top. Every 0.5 cm up along the arc represents a factor of 10 larger. (Top to bottom: R. Williams and the Hubble Deep Field Team [STScI] and

NASA; ESA/Hubble & NASA; NASA/JPL-Caltech/University of Wisconsin; NASA/SDO/HMI; NASA/NOAA/GSFC/Suomi NPP/VIIRS/Norman Kuring; Jose Luis Pelaez/Getty Images; Lee D. Simon/Mary Martin/Science Source; Courtesy of Florian Banhart/University of Mainz)

the Sun to the distance between Earth and the nearby stars.

This wide range of sizes underscores the fact that astronomy *synthesizes* or brings together information from many other fields of science. While we cannot go to the ends of the universe to examine all its components, the light from the universe coming to us, combined with our understanding of the laws of nature, provides invaluable insights into how the various components of the cosmos work and how they interact with each other. We will discuss some of the underlying principles of science as we need them.

What, then, have astronomers seen of the universe? **Figure 1-30** presents examples of the types of objects we will explore in this text. An increasing number of planets like Jupiter, rich in hydrogen and helium (Figure 1-30a), as well as rocky planets similar to Earth, are being discovered orbiting other stars. Much smaller pieces of space debris—some of rock and metal called **asteroids** or **meteoroids** (Figure 1-30b), and others of rock and ice called **comets** (Figure 1-30c)—orbit the Sun (Figure 1-30d) and

other stars. Vast stores of interstellar gas and dust are found in many galaxies; these “clouds” are often the incubators of new generations of stars (Figure 1-30e). Stars by the millions, billions, or even trillions are held together in galaxies by the force of gravity (Figure 1-30f). Most galaxies contain **black holes**, objects with such strong gravitational attraction that nothing can escape from them (Figure 1-30g). Groups of galaxies, called clusters, are held together by gravity (Figure 1-30h), and clusters of galaxies are grouped together in superclusters. Huge quantities of intergalactic gas are often found between galaxies (Figure 1-30i).

Every object in astronomy is constantly changing—each has an origin, an active period you might consider as its “life,” and each will have an end. In addition to examining the objects that fill the universe, we will also study the processes that cause them to change. After all is said and done, you will discover that all the matter and energy that astronomers have detected are but the tip of the cosmic iceberg—there is much more in the universe, but astronomers do not yet know its nature.

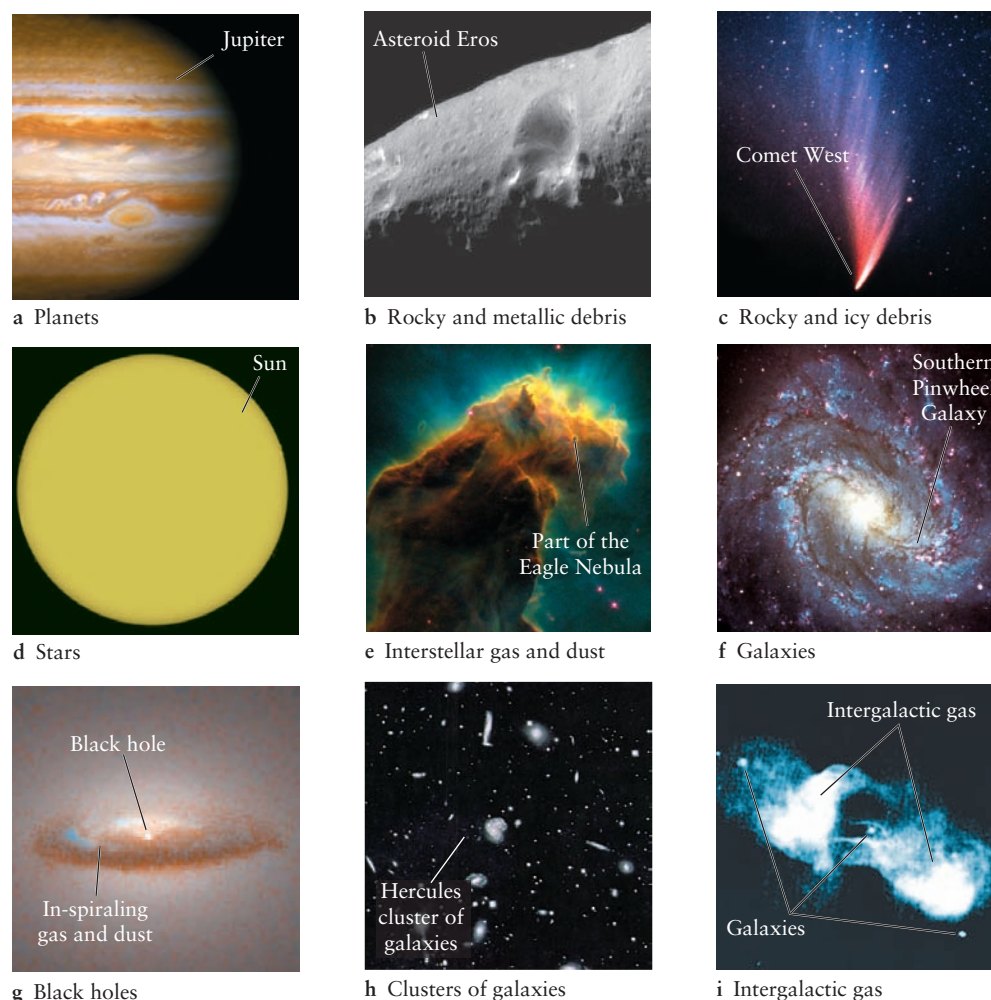


FIGURE 1-30 Inventory of the Universe Pictured here are examples of the major categories of objects that have been found throughout the universe. You will discover more about each type in the chapters that follow. (a: NASA/Hubblesite; b: NASA; c: Peter

Stättmayer/ESO; d: Big Bear Solar Observatory/New Jersey Institute of Technology; e: NASA/Jeff Hester & Paul Scowen; f: Australian Astronomical Observatory/David Malin Images; g: NASA; h: NOAO/AURA/NSF; i: Image courtesy of NRAO/N.F. Comins and F. N. Owen)

SUMMARY OF KEY IDEAS

Navigating the Night Sky

- The surface of the celestial sphere is divided into 88 unequal areas called constellations.
- The boundaries of the constellations run along lines of constant right ascension or declination, which are perpendicular to each other.

Earthly Cycles

- The celestial sphere appears to revolve around Earth once in each day-night cycle. In fact, it is Earth's rotation that causes this apparent motion.
- The poles and equator of the celestial sphere are determined by extending the axis of rotation and the equatorial plane of Earth out onto the celestial sphere.

The Seasons

- Earth's axis of rotation is tilted at an angle of $23\frac{1}{2}^\circ$ from a line perpendicular to the plane of Earth's orbit around the Sun (the plane of the ecliptic). This tilt causes the seasons.
- Equinoxes and solstices are significant points along Earth's orbit that are determined by the relationship between the Sun's path on the celestial sphere (the ecliptic) and the celestial equator.
- Earth's axis of rotation slowly changes direction relative to the stars over thousands of years, a phenomenon called precession. Precession is caused by the gravitational pull of the Sun and the Moon on Earth's equatorial bulge.
- The length of the day is based on Earth's rotation rate and the average motion of Earth around the Sun. These effects combine to produce the 24-hour day.

The Phases of the Moon

- The phases of the Moon are caused by the relative positions of Earth, the Moon, and the Sun. The Moon completes one cycle of phases in a synodic month, which averages $29\frac{1}{2}$ days.
- The Moon completes one orbit around Earth as measured with respect to the stars in a sidereal month, which averages 27.3 days.

Eclipses

- The shadow of an object has two parts: the umbra, where direct light from the source is completely blocked; and the penumbra, where the light source is only partially obscured.
- A lunar eclipse occurs when the Moon moves through Earth's shadow. During a lunar eclipse, the Sun, Earth, and the Moon are in alignment, with Earth between the Sun and the Moon, and the Moon is in the plane of the ecliptic.
- A solar eclipse occurs when a strip of Earth passes through the Moon's umbra or penumbra. During a

solar eclipse, the Sun, Earth, and the Moon are in alignment, with the Moon between Earth and the Sun, and the Moon is in the plane of the ecliptic.

- Depending on the relative positions of the Sun, the Moon, and Earth, lunar eclipses may be penumbral, partial, or total, and solar eclipses may be annular, partial, or total.

Scales of the Universe

- Astronomy examines objects that range in size from the parts of an atom ($\sim 10^{-15}$ m) to the size of the observable universe ($\sim 10^{26}$ m).
- Scientific notation is a convenient shorthand for writing very large and very small numbers.

WHAT DID YOU THINK?

1 *What do astronomers define as constellations?*

Astronomers sometimes use the common definition of a constellation as a pattern of stars. Formally, however, a constellation is an entire area of the celestial sphere and all the stars and other objects in it. Viewed from Earth, the entire sky is covered by 88 different-sized constellations. If there is any room for confusion, astronomers refer to the patterns as asterisms.

2 *Is the North Star—Polaris—the brightest star in the night sky?*

No. Polaris is a star of medium brightness compared with other stars visible to the naked eye.

3 *What causes the seasons?*

The tilt of Earth's rotation axis with respect to the ecliptic causes the seasons. They are not caused by the changing distance from Earth to the Sun that results from the oval shape of Earth's orbit.

4 *When is Earth closest to the Sun?*

Earth is closest to the Sun on or around January 3 of each year.

5 *How many zodiac constellations are there?*

There are 13 zodiac constellations, the least-known one being Ophiuchus.

6 *Does the Moon have a dark side that we never see from Earth?*

Half of the Moon is always dark. Whenever we see less than a full Moon, we are seeing part of the Moon's dark side. So, the dark side of the Moon is not the same as the far side of the Moon, which we never see from Earth.

7 *Is the Moon ever visible during the daytime?*

The Moon is visible at some time during daylight hours almost every day of the year. Different phases are visible during different times of the day.

8 *What causes lunar and solar eclipses?*

When the Moon is crossing the ecliptic in the full or new phase, the shadows of Earth or the Moon, respectively, then fall on the Moon or Earth. These shadows on the respective surfaces are eclipses.

Review Questions

- Where is the zenith?
 - directly overhead
 - along the celestial equator
 - on the boundary between land and sky
 - along the path that the Sun follows throughout the day
 - at the highest point the Sun passes above the celestial equator at any time during the year
- How many zodiac constellations are there?
 - 2
 - 12
 - 13
 - 56
 - 88
- Which of the following lies on the celestial sphere directly over Earth's equator?
 - ecliptic
 - celestial equator
 - north celestial pole
 - south celestial pole
 - horizon
- The length of time it takes Earth to orbit the Sun is:
 - an hour
 - a day
 - a month
 - a year
 - a century
- How are constellations useful to astronomers?
- What is the celestial sphere, and why is this ancient concept still useful today?
- What is the celestial equator, and how is it related to Earth's equator? How are the north and south celestial poles related to Earth's axis of rotation?
- What is the ecliptic, and why is it tilted with respect to the celestial equator?
- By about how many degrees does the Sun move along the ecliptic each day? Explain why.
- Through how many constellations does the Sun move every day? Explain why.
- Through how many constellations does the Sun move every year? Which ones are they?
- What are the vernal and autumnal equinoxes? What are the summer and winter solstices? How are these four events related to the ecliptic and the celestial equator?
- What is precession, and how does it affect our view of the heavens?
- How does the daily path of the Sun across the sky change with the seasons?
- Why is it warmer in the summer than in the winter?
- Why does the Moon exhibit phases?
- What is the difference between a sidereal month and a synodic month? Which is longer? Why?
- What is a penumbral eclipse of the Moon? Why is it easy to overlook such an eclipse?
- Which type of eclipse—lunar or solar—have more people seen? Why?

- How is an annular eclipse of the Sun different from a total eclipse of the Sun? What causes this difference?
- When is the next leap year?
- At which phase(s) of the Moon does a solar eclipse occur? A lunar eclipse?
- During what phase is the Moon “up” least in the daytime?

Got It?

- Describe the Orion constellation.
- In what season in the southern hemisphere is Earth closest to the Sun? In what season in the northern hemisphere is Earth closest to the Sun? If the changing distance from Earth to the Sun caused the seasons, what should be true about the answers to the first two parts of this question?
- What observational evidence do we have that the Moon does not make its own light?
- What is the difference between the “far” side of the Moon and the “dark” side of the Moon?
- Which of the following statements is correct for someone standing at one of Earth's poles?
 - The Sun rises and sets every day.
 - The Sun is directly overhead on the summer solstice in that hemisphere.
 - The Sun is in the sky continuously for 6 months.
 - The Moon is never visible.
 - The stars rise vertically and set vertically.

Advanced Questions

- During which phase(s) does the Moon rise after sunrise and before sunset? After sunset and before sunrise? At sunset? At sunrise?
- At what places on Earth is Polaris seen on the horizon?
- What is the phase of the Moon if it
 - rises at 3 A.M.?
 - sets at 9 P.M.?
 At what time does
 - the full Moon set?
 - the first quarter Moon rise?
- Why is a small crescent of light often observed on the Moon when it is exactly in the new phase?

Discussion Questions

- Examine the list of the 88 constellations in Appendix H: The Constellations. Are there any constellations whose names obviously date from modern times? Where are these constellations located? Why do they not have ancient names?
- In his novel *King Solomon's Mines*, H. Rider Haggard described a total solar eclipse that was seen in both South Africa and the British Isles. Is such an eclipse possible? Why or why not?



R I V I U X I G Gravity is the only universal force of attraction in the universe. Despite the magnificent solitude that astronaut Bruce McCandless II experienced floating a football-field length from the space shuttle *Challenger*, he was being held in orbit around Earth with almost the same amount of gravitational attraction that our planet has on each of us. He is falling toward Earth but continually missing it. Why? (STS-41B, NASA)

Gravitation and the Motion of the Planets

WHAT DO YOU THINK?

- 1 What makes a theory scientific?
- 2 What is the shape of Earth's orbit around the Sun?
- 3 Do the planets orbit the Sun at constant speeds?
- 4 Do all of the planets orbit the Sun at the same speed?
- 5 How does an object's mass differ when measured on Earth and on the Moon?
- 6 Do astronauts orbiting Earth feel the force of gravity from our planet?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

We begin this chapter by examining the nature of science, scientific investigation, and scientific inquiry. With the process of scientific investigation in mind, we explore how scientists came to discover that Earth and the other planets orbit the Sun. Then, in the spirit of scientific inquiry, we ask and answer the question “why do they have these orbits?”

In this chapter you will discover

- what makes a theory scientific
- the scientific discoveries that revealed that Earth is not at the center of the universe, as previously believed
- Copernicus’s argument that the planets orbit the Sun
- why the direction of motion of each planet on the celestial sphere sometimes changes
- that Kepler’s determination of the shapes and other properties of planetary orbits depended on the careful observations of his mentor Tycho Brahe
- how Isaac Newton formulated an equation to describe the force of gravity and how he thereby explained why the planets and moons remain in orbit

SCIENCE: KEY TO COMPREHENDING THE COSMOS

The groundwork for modern science was laid down by Greek “natural philosophers” beginning about 2500 years ago, when Pythagoras and his followers began using mathematics to describe natural phenomena. These Greek ideas, translated into Arabic, were rediscovered in the West in the seventeenth century. They led to the development of the scientific method of examining, understanding, and predicting how things work.

2-1 Science is both a body of knowledge and a process of learning about nature

Science is actually two related things. First, it is a body of knowledge that we acquire through observations, experiments, and mathematical calculations. The fact the planets orbit the Sun, along with equations that describe and predict these motions, are examples of that knowledge.

Second, science is a process for gaining more knowledge and deeper understanding of nature in a way that ensures that the information can be tested and thereby accepted by everyone. Science as a process is also called the **scientific method**, which describes how scientists ideally go about observing, explaining, and predicting physical reality.

The scientific method (Figure 2-1) can begin in a variety of places, but most often it starts with people making observations or doing experiments. Observations of planetary orbits provided to Johannes Kepler enabled him to derive Kepler’s laws, which are presented later in this chapter. These equations then accurately predicted the paths of other bodies in the solar system, such as moons orbiting planets. Understanding why Kepler’s laws are correct required understanding the force of gravity. English physicists Robert Hooke and Isaac Newton proposed that the force of gravity decreases inversely with the square of the distance between any two

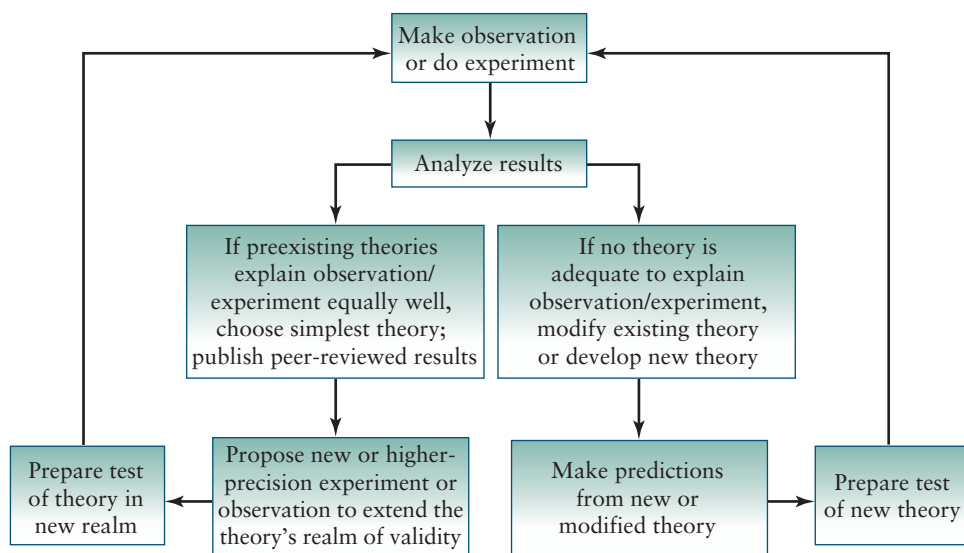


Figure 2-1 The Scientific Method This flowchart shows the basic steps in the process by which scientists study nature and develop new scientific theories. Different scientists start at different places on this chart, including making observations or doing experiments, creating or modifying scientific theories, or making predictions from theories. Anyone interested in some aspect of science and willing to learn the tools of the trade can participate in the adventure.

objects, such as Earth and the Moon. (This means, for example, that if you double the distance, the force of gravity is four times weaker.) Newton went on to derive Kepler's laws based on this assumption about the force of gravity. The force of gravity was thus added to the body of scientific knowledge and was then used to explain existing observations and to make predictions about previously unobserved motions.

If a theory exists that purports to explain previous observations, new observations or experimental results are then compared with the predictions of the theory. If the new data and old theory are not consistent, then a *hypothesis* that modifies or replaces the existing explanation is proposed. (If no theory explains observations or experimental results, a new hypothesis is proposed to explain them.) Hypotheses on related topics that make accurate predictions are incorporated together as a **scientific theory** (often just called a **theory**). Scientific theories are presented in the form of mathematical equations and accompanying explanations.

An interesting example of beginning a scientific inquiry with equations, rather than observations or experiments, is the discovery of a particle called the Higgs boson. This particle is what causes matter to have mass, which in turn allows matter to interact gravitationally, among other things. In 1964, Scottish physicist Peter Higgs, along with other scientists, predicted the existence of this particle based on powerful equations that describe the properties of matter. Based on this prediction, experiments were run to discover the Higgs boson. Scientists believe that it was observed in July 2012, at the Large Hadron Collider in Switzerland.

In everyday conversation, a theory is an idea based on common sense, intuition, or deep-seated personal beliefs. Such theories neither originate in equations nor do they usually lead to rigorous predictions. The word *theory* in science has a very different meaning. A scientific theory is an explanation of observations or experimental results that can be described quantitatively (that is, in terms of equations) and tested formally. The mathematical description used in a scientific theory is considered a **model** of the real system. For example, Newton's *theory* (or, in earlier usage, *law*) of gravitation is written as an equation that predicts how bodies attract each other. (The word *gravity* is often used as shorthand for *gravitation*, and both are used in this book.)

1 As just noted, to be considered scientific, a theory must make *testable* predictions that can be verified by making new observations or doing new experiments. Testing is a crucial aspect of the scientific method, which also requires that the theory accurately forecast the results of new observations in its realm of validity. Newton's law of gravitation predicts that the Sun's gravitational force makes the planets move in elliptical orbits, and it predicts how long it should take each planet to orbit the

Sun. As we will see shortly, observations have confirmed most of these predictions.

For a theory to be considered scientific, it must also be potentially possible to disprove it. For example, Newton's law of gravitation can be tested and potentially disproven by observations and thus qualifies as a scientific theory. The idea that Earth was created in 6 days cannot be tested, much less disproved. It is not a scientific theory but rather a matter of faith.

One important theme in science is to look for patterns that allow seemingly unrelated events or activities to be explained by one theory. For example, Newton observed that the Moon's motion around Earth had the same behavior as the motion of a flying cannonball. Indeed, if you fired a cannonball fast enough, it would orbit Earth just as the Moon does. He hypothesized that they were both responding to Earth's gravitational attraction, which led to his successfully applying the same equations to describe their motions. The motion of the planets around the Sun, he found, could also be described by the same equations. One theory. Three applications. Very satisfying! It is worth noting that Newton's theory of gravity has since been applied successfully to myriad other situations.

Science strives to explain as many things as possible with as few theories as possible. As another example, we see billions upon billions of objects in the universe. It would be virtually impossible to study all of them separately so that we could come up with detailed descriptions of each one. Fortunately, individual theories explaining each object are not necessary. Scientists overcome this problem by noting that many of the bodies in space appear similar to each other. By categorizing them suitably and then applying the scientific method to these groups of objects, we form a few theories that describe many objects and how they have evolved. These few theories can then be tested and refined as necessary. Such groupings of objects have proven invaluable, and they give us insights into the structure and organization of billions of stars and galaxies that are, indeed, very similar to one another.

Often several competing theories describe the same concepts with the same accuracy. In such cases, scientists choose the simplest one—namely, the theory that contains the fewest unproven assumptions. That basic tenet, formally expressed by the philosopher and Franciscan friar William of Occam in the fourteenth century, is known as **Occam's razor**. Indeed, the Sun-centered cosmology as refined by Johannes Kepler, which we are about to explore, was appealing because it made the same predictions within a simpler model than did the earlier Earth-centered cosmology. *Remember Occam's razor.*

Focus Question 2-1

Can you give an example of one scientific hypothesis and one nonscientific hypothesis not mentioned so far?

Insight Into Science

Science Is Inclusive In principle, a scientific theory can be created, modified, or tested by anyone inclined to do so. In practice, however, being involved in the scientific enterprise requires that you understand the mathematical tools of science. Assuring that theories are written in terms of equations so that they can be carefully analyzed and tested by others is part of the process intended to prevent the scientific method from being derailed.

If the predictions of a theory are inconsistent with observations, the theory is modified, applied in more limited circumstances, or discarded in favor of a more accurate explanation. For example, Newton's law of gravitation is entirely adequate for describing the motion of an apple falling to Earth, the flight of a soccer ball, or the path of Earth orbiting the Sun; however, it is inaccurate in describing the orbit of Mercury around the Sun or the behavior of matter in the vicinity of very dense concentrations of matter, like black holes. In these cases, Newton's law of gravitation is replaced by Einstein's theory of general relativity, which describes gravitational behavior more accurately and over a much wider range of conditions than Newton's law but at the cost of much greater mathematical complexity. (The Global Positioning System, or GPS, requires such accurate measurements that scientists must include general relativity in the equations used to determine the locations of GPS satellites orbiting Earth.)

The scientific method can be summarized in six words: *observe, hypothesize, predict, test, modify,*

Insight Into Science

Theories and Beliefs New theories are personal creations, but science is not a personal belief system. As stated in the previous Insight Into Science, scientific theories make predictions that can be tested by independent scientists. If everyone who performs tests of the theory's predictions gets results consistent with the theory, the theory is considered valid in that realm. In comparison, belief systems—such as which sports team or political system is best—are personal matters. People will always hold differing opinions about such issues.

economize. I urge you to watch for applications of the scientific method throughout this book. Our first encounter with it is the discovery that Earth orbits the Sun.

CHANGING OUR EARTH-CENTERED VIEW OF THE UNIVERSE

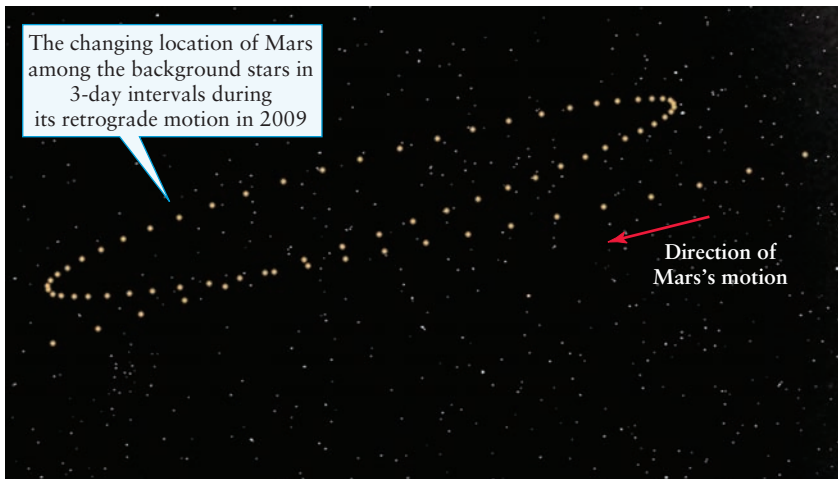
Early Greek astronomers tried to explain the motion of the five then-known planets: Mercury, Venus, Mars, Jupiter, and Saturn. Most people at that time held a *geocentric* view of the universe: Based on the observed motion of the celestial sphere, they believed that the Sun, the Moon, the stars, and the planets revolved around Earth. A theory of the overall structure and evolution of the universe is called a **cosmology**, so the prevailing Earth-centered cosmology was called *geocentric*. Geocentric cosmology, consistent as it is with casual observation of the sky, held sway until the sixteenth century, when Copernicus revolutionized our understanding of the cosmos.

2-2 The belief in a Sun-centered cosmology formed slowly

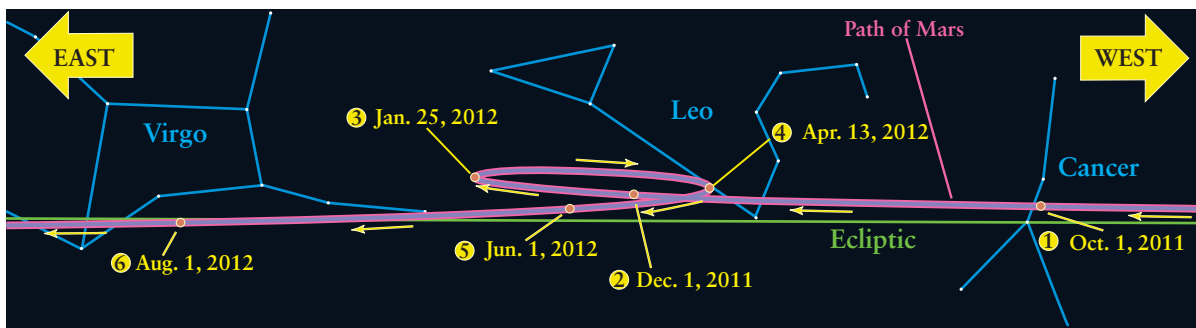
Explaining the motions of the five planets in a geocentric universe was one of the main challenges facing the astronomers of antiquity. The Greeks knew that the positions of the planets slowly shift relative to the “fixed” stars in the constellations. In fact, the word *planet* comes from a Greek term meaning “wanderer.” They also observed that planets do not move at uniform rates through the constellations. From night to night, as viewed in the northern hemisphere, the planets usually move slowly to the left (eastward) relative to the background stars. This movement is called **direct motion** or **prograde motion**. Occasionally, however, a planet seems to stop and then back up for several weeks or months. This reverse movement (to the right or westward relative to the background stars) is called **retrograde motion**. Both direct and retrograde motions are best observed by photographing (Figure 2-2a) or plotting (Figure 2-2b and 2-2c) the nightly position of a planet against the background stars over a long period.

All planetary motions on the celestial sphere are much slower than the apparent daily movement of the entire sky caused by Earth's rotation. Throughout a single night it is very difficult to detect motion of the planets among the stars. Therefore, the planets always rise with the stars in the eastern half of the sky and set with them in the western half.

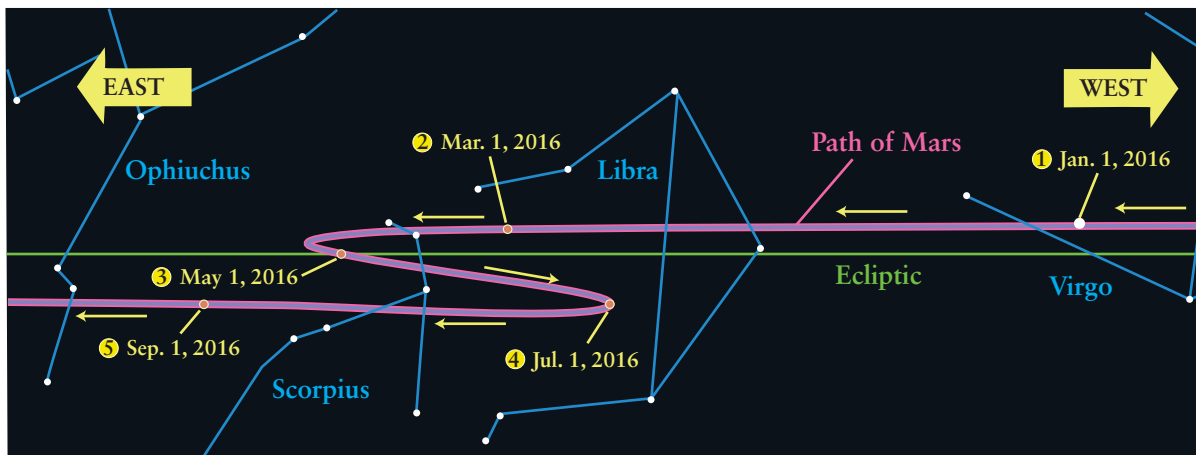
The effort to understand planetary motion—and especially to explain retrograde motion—in a geocentric



a



b



c

Figure 2-2 Paths of Mars (a) The retrograde motion of Mars as it would be seen in a series of images taken on the same photographic plate. (b) To help visualize this motion on the celestial sphere, astronomers often plot the position of Mars (or other body in retrograde motion) on a star chart. The

retrograde path is sometimes a loop north, as shown in (a) and (b), or south of the normal path, and sometimes an S-shaped path across the ecliptic. (c) In the middle of 2016, Mars will undergo an S-shaped retrograde motion.

cosmology resulted in an increasingly contrived and complex model (see **Discovery 2-1: Earth-Centered Universe**). The ancient Greek astronomer Aristarchus of Samos proposed a more straightforward explanation of planetary motion, namely, that all of the planets,

including Earth, revolve around the Sun. The retrograde motion of Mars in this **heliocentric** (Sun-centered) **cosmology** occurs because the faster-moving Earth overtakes and passes

Focus Question 2-2

Referring to Figure 2-2c, why do you think Mars is seen sometimes above the ecliptic and sometimes below it?

DISCOVERY 2-1

Earth-Centered Universe

Most of us find it hard to understand why anyone would believe that the Sun, planets, and stars orbit Earth. After all, we *know* that Earth spins on its axis. We *know* that the gravitational force from the Sun holds the planets in orbit, just as Earth's gravitational force holds the Moon in orbit. In childhood these facts become part of our understanding of the motions of the heavenly bodies.

Psychologists call this background information that we use to help explain things a *conceptual framework*. A conceptual framework contains all of the information we take for granted. For example, when the Sun rises, moves across the sky, and sets, we take for granted that it is Earth's rotation that causes the Sun's apparent motion.

Because they did not feel Earth move under their feet, or see any other indication that Earth is in motion, it did not even occur to our distant forebears to consider the possibility that we are in orbit around the Sun. The obvious conclusion for one who has a prescientific conceptual framework, even today, is that Earth stays put while objects in the heavens move around it. This prescientific conceptual framework for understanding the motions of the heavenly bodies was based on the senses and on common sense. That is, people observed motions and drew "obvious," commonsense conclusions. Today, we incorporate the known and tested laws of physics in our understanding of the natural world. Many of these realities are utterly counterintuitive and, therefore, the conceptual frameworks that we possess are less consistent with common sense

than those held in the past. Studying science helps us to develop intuition that is consistent with the actual workings of nature.

Geocentric Explanation of the Planets' Retrograde Motion The early Greeks, working in what we now know is an incorrect geocentric (Earth-centered) cosmology, developed many theories to account for the occasional retrograde motion of the planets and the resulting loops that the planets trace out against the background stars. One of the most successful (albeit incorrect) ideas was expounded by the last of the great ancient Greek astronomers, Ptolemy, who lived in Alexandria, Egypt. His basic concepts are sketched in **Figure D2-1**. Each planet is assumed to move in a small circle called an *epicycle*, the center of which moves in a larger circle called a *deferent*, whose center is offset from Earth. As viewed from Earth, the epicycle moves eastward along the deferent, and both the epicycle and the planet on it revolve in the same direction (counterclockwise in the figure).

Most of the time, the motion of the planet on its epicycle adds to the eastward motion of the epicycle on the deferent. Thus, the planet is seen from Earth to be in direct motion (to the left or eastward) against the background stars throughout most of the year (Figure D2-1a). However, when the planet is on the part of its epicycle nearest Earth, its motion along the epicycle subtracts from the motion of the epicycle along the deferent. The planet thus appears to slow and then halt its usual movement to the left (eastward motion) among the constellations and then

the Red Planet (**Figure 2-3**). The occasional retrograde movement of a planet is merely the result of our changing viewpoint as we orbit the Sun—an idea that is beautifully simple compared to the geocentric system with all of its complex planetary motions. (The word *heliocentric* is misleading. Although the local planets, moons, and small pieces of space debris do orbit the Sun, the stars and innumerable other objects in space do not. In fact, the Sun and the bodies that orbit it all orbit the center of our Milky Way Galaxy.)

2-3 Copernicus devised the first comprehensive heliocentric cosmology

Over the centuries following Aristarchus, increasingly accurate observations of the planets' locations revealed errors in the predictions of geocentric cosmology. To reconcile that cosmology with the data, more and more

complex motions were attributed to the planets. By the mid-1500s, the geocentric cosmology had become truly unwieldy in its efforts to predict the motions of the planets accurately. It was then that the Polish mathematician, lawyer, physician, economist, cleric, and artist Nicolaus Copernicus resurrected Aristarchus's theory. Copernicus was motivated by an effort to simplify the celestial scheme, and eventually his heliocentric cosmology did just that.

Pursuing the heliocentric cosmology, Copernicus determined, by observations, which planets are closer to the Sun than Earth and which are farther away. Because Mercury and Venus are always observed fairly near the Sun, he correctly concluded that their orbits must lie inside Earth's orbit. The other planets visible to Copernicus—Mars, Jupiter, and Saturn—can sometimes be seen high in the sky in the middle of the night, when the Sun is far below the horizon. This placement can occur only

seems to move to the right (westward) among the stars for a few weeks or months (Figure D2-1b). This concept of epicycles and deferents enabled Greek astronomers to explain the retrograde loops of the planets. Ptolemy's

geocentric cosmology endured as a useful predictor of the motions of the Sun, Moon, and planets for more than 1000 years.

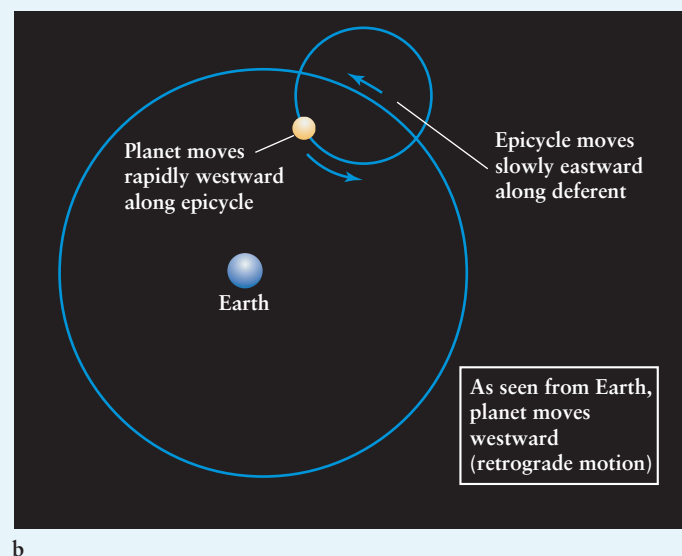
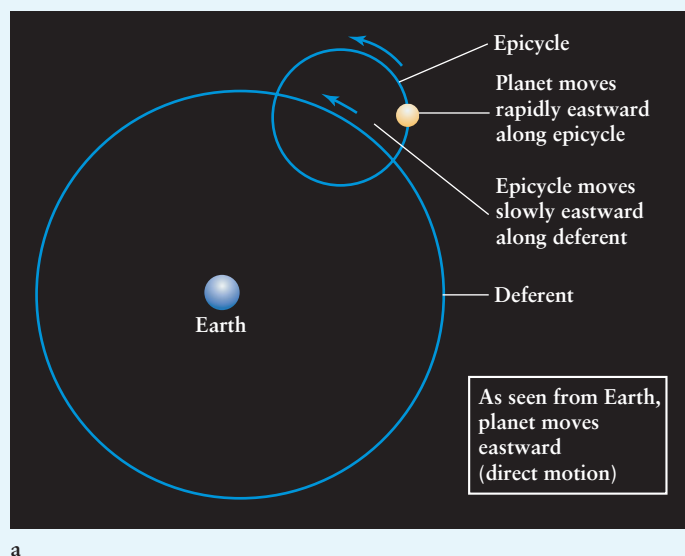


FIGURE D2-1 A Geocentric Explanation of Planetary Motion Each planet revolves around an epicycle, which in turn revolves around a deferent centered approximately on Earth. As seen from Earth, the speed of the planet on the epicycle

alternately (a) adds to or (b) subtracts from the speed of the epicycle on the deferent, thus producing alternating periods of direct and retrograde motions.

if Earth comes between the Sun and a planet. Copernicus therefore concluded (also correctly) that the orbits of Mars, Jupiter, and Saturn lie outside Earth's orbit. Today, Mercury and Venus are called **inferior planets**, while all the planets farther from the Sun than Earth are called **superior planets**.

The geometric arrangements among Earth, another planet, and the Sun are called **configurations**. For example, when Mercury or Venus is directly between Earth and the Sun (Figure 2-4), we say the planet is in a configuration called an **inferior conjunction**; when either of these planets is on the opposite side of the Sun from Earth, its configuration is called a **superior conjunction**.

The angle between the Sun and a planet as viewed from Earth is called the planet's **elongation**. A planet's elongation varies from zero degrees to a maximum value, depending upon where we see it in its orbit around the Sun. At *greatest eastern* or *greatest western elongation*,

Mercury and Venus are as far from the Sun in angle as they can be. This position is about 28° for Mercury and about 47° for Venus. The superior planets all have maximum elongations of 180° . When either Mercury or Venus rises before the Sun, the planet is visible in the eastern sky as a bright "star" and is often called the "morning star." Similarly, when either of these two planets sets after the Sun, the planet is visible in the western sky and is then called the "evening star." Because these two planets are not always at their greatest elongations, they are often very close in angle to the Sun. This positioning is especially true of Mercury, often making it



Nicolaus Copernicus (1473–1543)
(Universal History Archive/Getty Images)

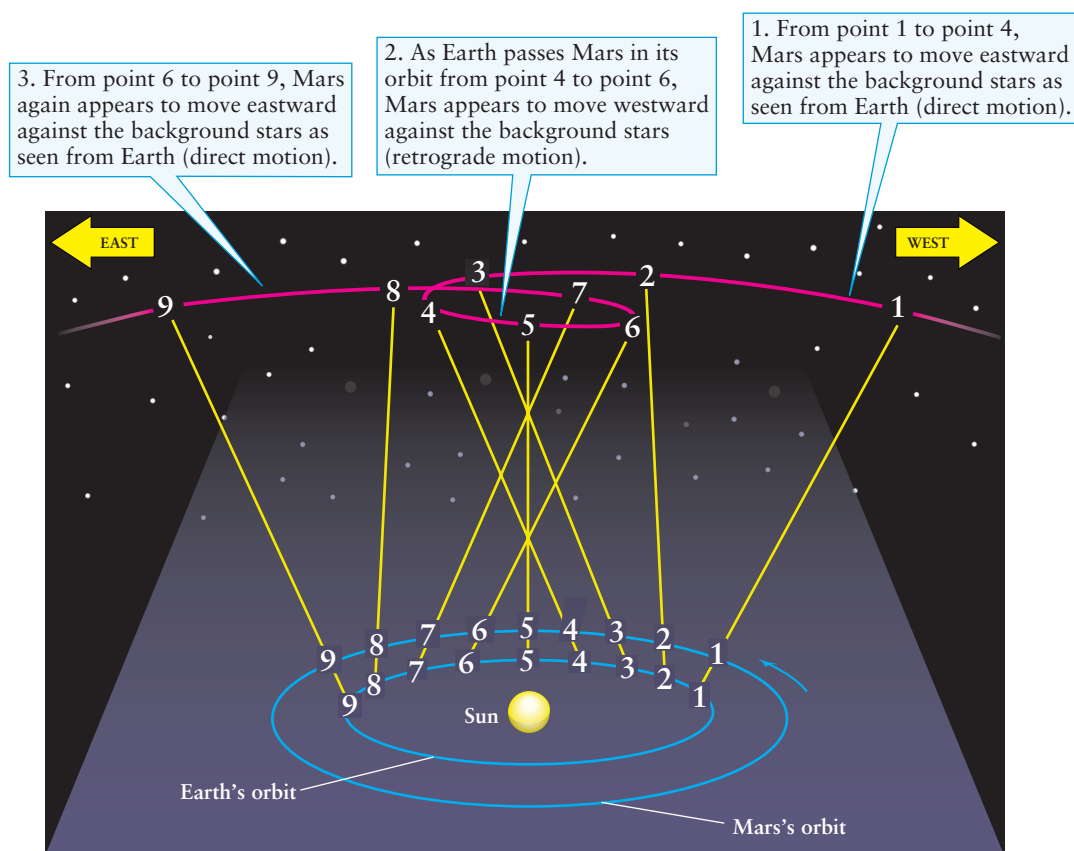


FIGURE 2-3 A Heliocentric Explanation of Retrograde Motion Earth travels around the Sun more rapidly than does Mars. Consequently, as Earth overtakes and passes this slower-

moving planet, Mars appears (from points 4 through 6) to move backward among the background stars for a few months.

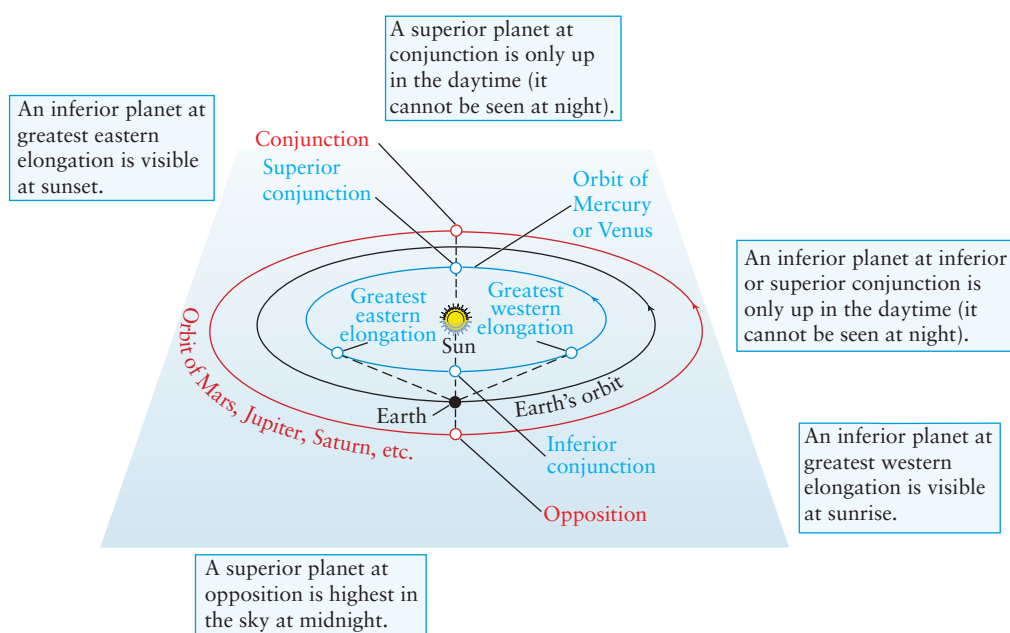


FIGURE 2-4 Planetary Configurations Key points along a planet's orbit have names, as shown. These points identify spe-

cific geometric arrangements between Earth, another planet, and the Sun.

hard to see from Earth. Venus is often nearly halfway up the sky at sunrise or sunset and therefore quite noticeable during much of its orbit. Because they are so bright and sometimes appear to change color due to the motion of Earth's atmosphere, Venus and Mercury are often mistaken for UFOs. (The same motion of the air causes the road in front of your car to shimmer on a hot day.)

Planets farther from the Sun than Earth have different configurations. When one of these planets is located behind the Sun, as seen from Earth, that planet is said to be in **conjunction**. When a planet is opposite the Sun in the sky, that planet is at **opposition**. It is not difficult to determine when a planet happens to be located at one of the key positions in Figure 2-4. For example, when Mars is at opposition, it appears high and bright in the sky at midnight.

It is relatively easy to follow a planet as it moves from one configuration to another. However, these observations alone do not tell us the planet's actual orbit, because Earth, from which we make the observations, is also moving. Copernicus was therefore careful to distinguish between two characteristic time intervals, or *periods*, of each planet.

Recall from Chapter 1 that a planet's **sidereal period** is the time it takes that body to make one complete orbit of the Sun as measured with respect to the distant stars; an observer fixed at the Sun's location watching that planet would see it start at one point on the celestial sphere, go around the sphere, and end at the same place it started from. *The sidereal period or orbit is the length of a year for each planet.*

The other useful time interval that Copernicus used is the **synodic period**. The synodic period is the time that elapses between two successive identical configurations as seen from Earth. This period can be from one opposition to the next, for example, or from one conjunction to

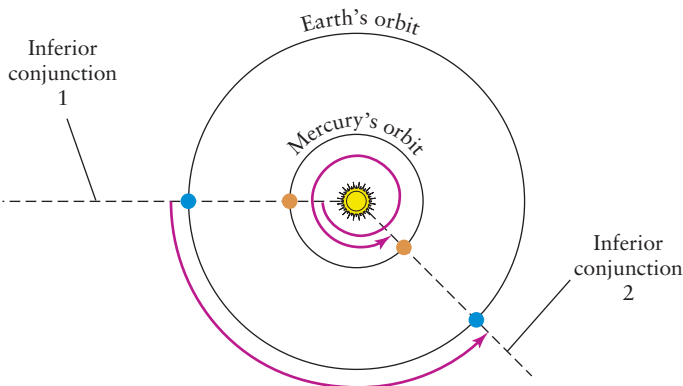


FIGURE 2-5 Synodic Period The time between consecutive conjunctions of Earth and Mercury is 116 days. As is typical of synodic periods for all planets, the location of Earth is different at the beginning and end of the period. You can visualize the synodic periods of the outer planets by putting Earth in Mercury's place in this figure and putting one of the outer planets in Earth's place.

TABLE 2-1 SYNODIC AND SIDEREAL PERIODS OF THE PLANETS (IN EARTH YEARS)

	Synodic (year)	Sidereal (year)
Mercury	0.318	0.241
Venus	1.599	0.616
Earth	—	1.0
Mars	2.136	1.9
Jupiter	1.092	11.9
Saturn	1.035	29.5
Uranus	1.013	84.0
Neptune	1.008	164.8

the next (of the same type, Figure 2-5). The synodic period tells us, among other things, when to expect a planet to be closest to Earth and, therefore, most easily studied.

Focus Question 2-3
Why do I say “of the same type” in this paragraph?

Thus, nearly 500 years ago, Copernicus was able to obtain the first six entries shown in Table 2-1 (the others are contemporary results included for completeness). Copernicus was then able to devise a straightforward geometric method for determining the distances of the planets from the Sun. His answers turned out to be remarkably close to the modern values, as shown in Table 2-2. From these two tables it is apparent that the farther a planet is from the Sun, the longer it takes to complete its orbit.

Copernicus presented his heliocentric cosmology, including supporting observations and calculations, in a book entitled *De revolutionibus orbium coelestium* (*On the Revolutions of the Celestial Spheres*), which was published in 1543, the year of his death. His great

TABLE 2-2 AVERAGE DISTANCES OF THE PLANETS FROM THE SUN

	Measurement (AU)	
	By Copernicus	Modern
Mercury	0.38	0.39
Venus	0.72	0.72
Earth	1.00	1.00
Mars	1.52	1.52
Jupiter	5.22	5.20
Saturn	9.07	9.54
Uranus	Unknown	19.19
Neptune	Unknown	30.06

insight was the conceptual simplicity of a heliocentric cosmology compared to geocentric views, especially in explaining retrograde motion. However, Copernicus incorrectly assumed that the planets travel along circular paths around the Sun. Without using epicycles similar to those used in geocentric theory (see Figure D2-1), many of his predictions were no more accurate than those of the earlier theory, as we will see shortly.

2-4 Tycho Brahe made astronomical observations that disproved ancient ideas about the heavens

In November 1572, a bright star suddenly appeared in the constellation Cassiopeia. At first, it was even brighter than Venus, but then it began to grow dim. After 18 months the star faded from view.

Modern astronomers recognize this event as a supernova explosion, the violent death of a certain type of star (see Chapter 11). In the sixteenth century, however, the prevailing opinion was quite different. Teachings dating back to Aristotle and Plato argued that the heavens were permanent and unalterable. From that perspective, the “new star” of 1572 could not really be a star at all, because the heavens do not change. Many astronomers and theologians of the day argued that the sighting must be some sort of bright object quite near Earth, perhaps not much farther away than the clouds overhead. A 25-year-old Danish astronomer named Tycho Brahe realized that straightforward observations might reveal the distance to this object.

Consider what happens when two people look at a nearby object from different places—they see it in different positions relative to the things behind it. Furthermore, their heads face at different angles when looking at it. This variation in angle that occurs when viewing a nearby object from different locations is called **parallax** (Figure 2-6).

Tycho reasoned as follows: If the new star is nearby, its position should shift against the background stars over the course of a night (Figure 2-7a). His

careful observations, done in the spirit of the scientific method, failed to reveal any parallax, and so the new star had to be far away, farther from Earth than anyone had imagined (Figure 2-7b). Tycho summarized his findings in a small book, *De nova stella* (*On the New Star*), published in 1573.

From 1576 to 1597, Tycho made comprehensive observations, measuring planetary positions with an accuracy of 1 arcminute (arcmin; see Section 1-5), about as precise as is possible with nontelescopic instruments (the telescope was invented in 1608). Upon Tycho’s death in 1601, most of his invaluable astronomical records were given to his gifted assistant, Johannes Kepler.



Tycho Brahe (1546–1601), also depicted within the portrait of Kepler (INTERFOTO/Alamy)

Focus Question 2-4

Describe a simple experiment to demonstrate that your eyes (and, implicitly, your brain) use parallax to determine distances.

Insight Into Science

Take a Fresh Look When a scientific concept is hard to visualize, try another perspective. For example, a planet’s sidereal period of orbit is easy to understand when viewed from the Sun but more complicated as seen from Earth. The synodic period of each planet, on the other hand, is easily determined from Earth. As we will see, especially when we study Einstein’s theories of relativity, each of these perspectives is called a *frame of reference*.

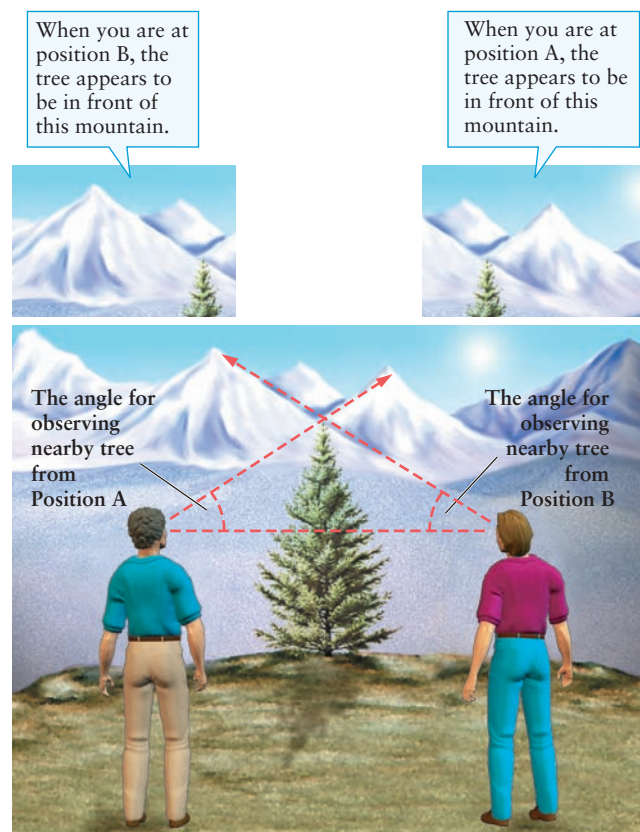


FIGURE 2-6 Parallax Nearby objects are viewed at different angles from different places, an effect called parallax. These objects also appear to be in different places with respect to more distant objects when viewed by observers located at different positions. Parallax is used by astronomers, surveyors, and sailors to determine distances.

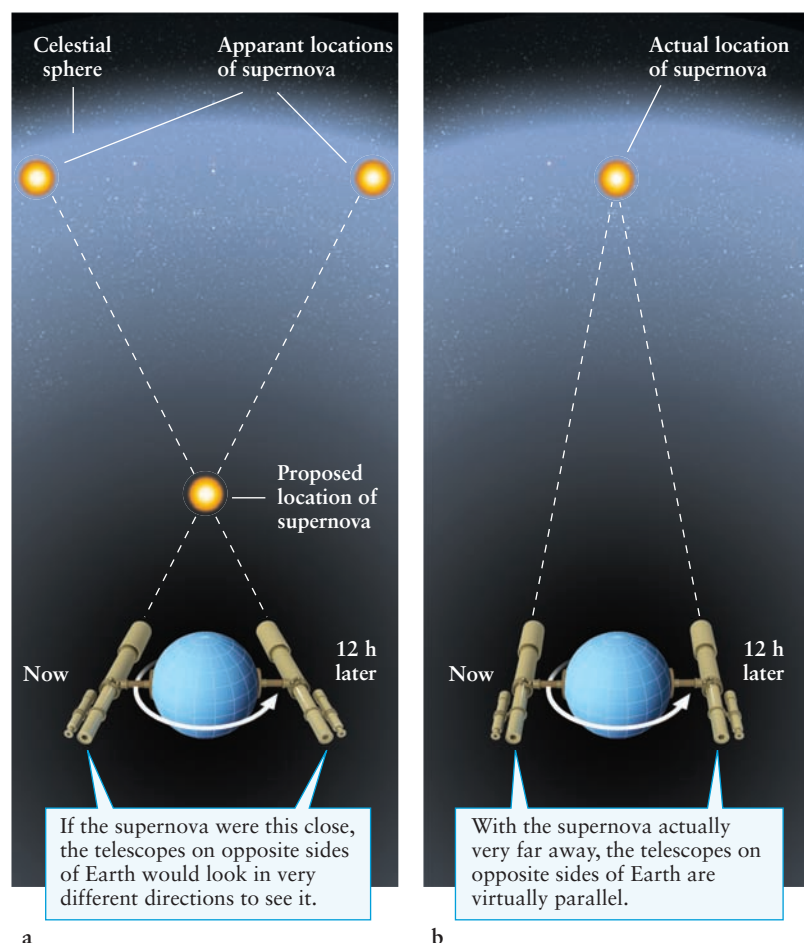


FIGURE 2-7 The Parallax of a Nearby Object in Space Tycho thought that Earth does not rotate and that the stars revolve around it. From our modern perspective, the changing position of the supernova would be due to Earth's rotation as shown. **(a)** Tycho argued that if an object is near Earth, its position relative to the background stars should change over the course of a night. **(b)** Tycho failed to measure such changes for the supernova in 1572. This is illustrated in (b) by the two telescopes being parallel to each other. He, therefore, concluded that the object was far from Earth.

KEPLER'S AND NEWTON'S LAWS

Until Kepler's time, astronomers had assumed that heavenly objects move in circles. For philosophical and aesthetic reasons, circles were considered the most perfect and most harmonious of all geometric shapes. However, using circular orbits failed to yield accurate predictions for the positions of the planets. For years, Kepler tried to find a shape for orbits that would fit Tycho's observations of the planets' positions against the background of distant stars. Finally, he began working with a geometric form called an **ellipse**.

2-5 Kepler's laws describe orbital shapes, changing speeds, and the lengths of planetary years

You can draw an ellipse as shown in Figure 2-8a. Each thumbtack is at a **focus** (plural **foci**). The longest diameter (major axis) across an ellipse passes through both foci. Half of that distance is called the **semimajor axis**. In astronomy, the length of the semimajor axis is also the average distance between a planet and the Sun.

2 To Kepler's delight, the ellipse turned out to be the curve for which he had been searching. Predictions of the locations of planets based on elliptical paths were

in very close agreement with where the planets actually were. Keep in mind that the following three laws that Kepler discovered merely quantified (gave equations for) observations that Tycho had made—Kepler did not have a theory to explain them. That would come nearly 80 years later with the work of Isaac Newton.

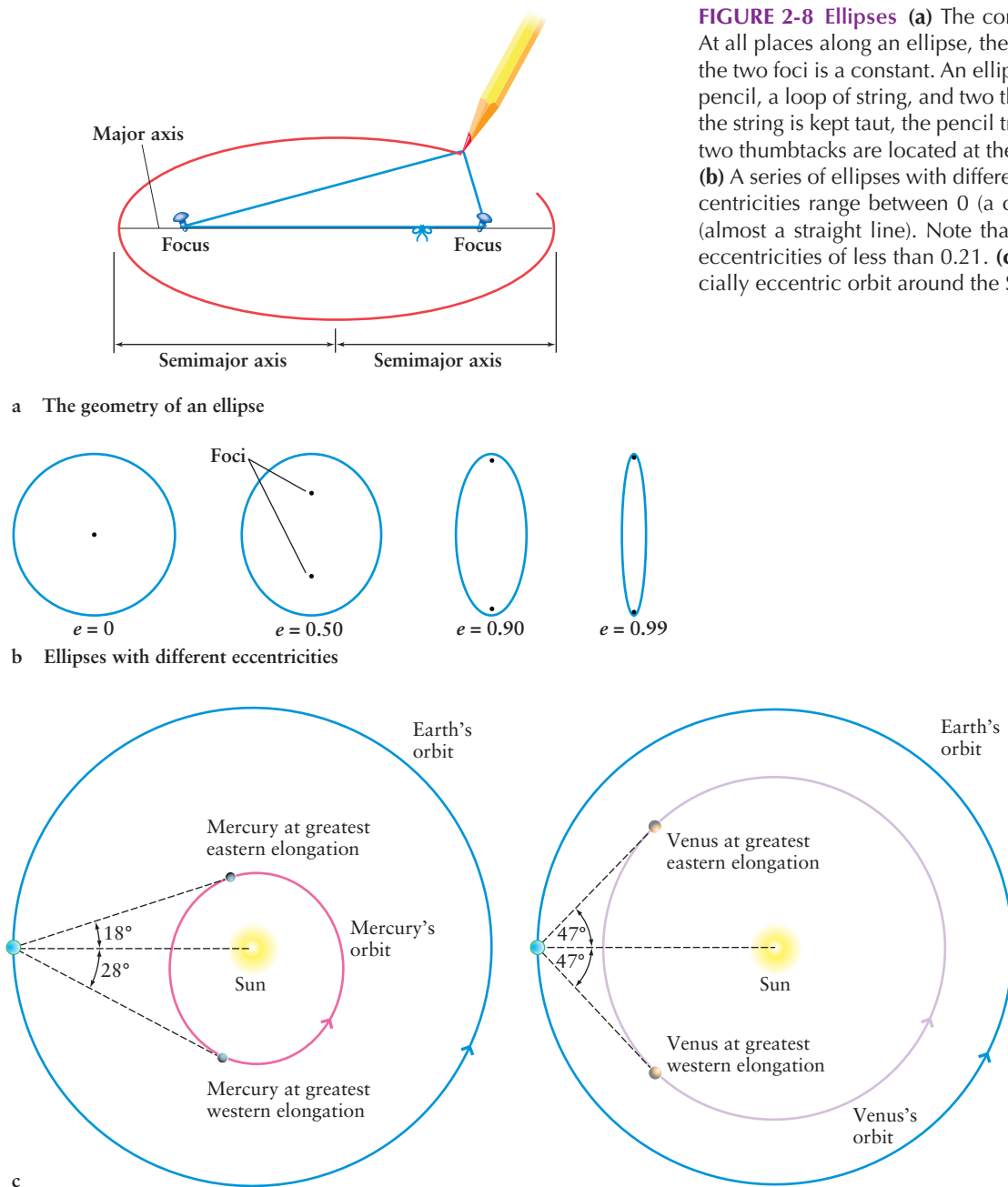
Kepler published his discovery of elliptical orbits in 1609 in a book known today as *New Astronomy*. This important discovery is now considered the first of **Kepler's laws**:

Kepler's First Law: *The orbit of each planet around the Sun is an ellipse with the Sun at one focus.*

The shapes of ellipses have two extremes. The roundest ellipse, occurring when the two foci merge, is a circle. The most elongated ellipse is nearly a straight line. The shape of a planet's orbit around the Sun is described by its *orbital eccentricity*, designated by the letter e , which ranges from 0 (a circular orbit) to just under 1.0 (nearly



Johannes Kepler
(1571–1630) (Science Source)



a straight line). Figure 2-8b shows a sequence of ellipses and their associated eccentricities. Observations have revealed that there is no object at the second focus of each elliptical planetary orbit. Figure 2-8c shows the effect of orbital eccentricity. For example, the maximum elongation of Mercury ($e = 0.21$) seen from Earth varies by 10° , while the maximum elongation of Venus ($e = 0.01$) varies by less than 1° .

3 Tycho's observations also showed Kepler that planets do not move at uniform speeds along their orbits. Rather, a planet moves fastest when it is nearest the Sun, a point on its orbit called **perihelion**. Conversely,

a planet moves most slowly when it is farthest from the Sun, a point called its **aphelion**.

After much trial and error, Kepler discovered a way to describe how fast a planet moves anywhere along its orbit. This discovery, also published in *New Astronomy*, is illustrated in **Figure 2-9**. Suppose that it takes 30 days for a planet to go from point A to point B. During that time, the line joining the Sun and the planet sweeps out a nearly triangular area (shaded in Figure 2-9). Kepler discovered that the line joining the Sun and the planet sweeps out the same area during any other 30-day interval. In other words, if the planet

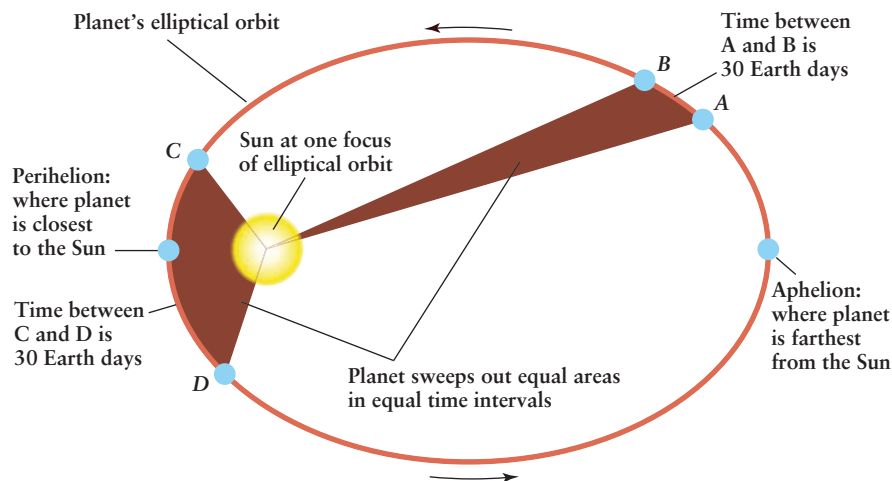


FIGURE 2-9 Kepler's First and Second Laws According to Kepler's first law, every planet travels around the Sun along an elliptical orbit with the Sun at one focus. According to his second law, the line joining the planet and the Sun sweeps out equal areas (the burgundy-colored regions) in equal intervals of time (time from A to B equals time from C to D). *Note:* This drawing shows a highly elliptical orbit, with $e = 0.74$. Even though this is a much greater eccentricity than that of any planet in the solar system, the concept still applies to all planets and other orbiting bodies.

also takes 30 days to go from point C to point D, then the two shaded segments in Figure 2-9 are equal in area. Kepler's second law, also called the **law of equal areas**, can be stated thus:

Kepler's Second Law: *A line joining a planet and the Sun sweeps out equal areas in equal intervals of time.*

A consequence of Kepler's second law is that each planet's speed decreases as it moves from perihelion to aphelion. The speed then increases as the planet moves from aphelion toward perihelion.

Kepler was also able to relate a planet's year to its distance from the Sun. This discovery, published in 1619, is Kepler's third law. This relationship predicts the planet's sidereal period if we know the length of the semimajor axis of the planet's orbit:

Kepler's Third Law: *The square of a planet's sidereal period around the Sun is directly proportional to the cube of the length of its orbit's semimajor axis.*

The relationship is easiest to use if we let P represent the sidereal period of an object's orbit around the Sun in Earth years and a represent the length of its semimajor

axis (that is, its average distance from the Sun) measured in **astronomical units (AU)**. One astronomical unit is the average distance from Earth to the Sun; hence, $a = 1$ for Earth. The astronomical unit is used when measuring distances between objects in the solar system, because no powers of ten are needed, as they would be if these distances were referred to in kilometers or miles. (See **Discovery 2-2: Units of Astronomical Distance** for more details.) Now we can write Kepler's third law for all objects orbiting the Sun as

$$P^2 = a^3$$

4 This equation says that a planet closer to the Sun has a shorter year than does a planet farther from the Sun. Using this equation with Kepler's second law reveals that planets closer to the Sun move more rapidly than those farther away. Using data from Table 2-1 and Table 2-2, we can demonstrate Kepler's third law as shown in **Table 2-3**.

Focus Question 2-5

What is the shape of the International Space Station's orbit around Earth?

Focus Question 2-6

We saw in Chapter 1 that the Moon's orbit around Earth is not circular. Where in its orbit is the Moon moving fastest, and where is it moving slowest?

TABLE 2-3 A DEMONSTRATION OF KEPLER'S THIRD LAW

	Sidereal period P (year)	Semimajor axis a (AU)	P^2	=	a^3
Mercury	0.24	0.39	0.06		0.06
Venus	0.61	0.72	0.37		0.37
Earth	1.00	1.00	1.00		1.00
Mars	1.88	1.52	3.53		3.51
Jupiter	11.86	5.20	140.7		140.6
Saturn	29.46	9.54	867.9		868.3
Uranus	84.01	19.19	7058		7067
Neptune	164.79	30.06	27,160		27,160

DISCOVERY 2-2

UNITS OF ASTRONOMICAL DISTANCE

Throughout this book we will find that some of our traditional units of measure become cumbersome. It is fine to use kilometers to measure the diameters of craters on the Moon or the heights of volcanoes on Mars. However, it is as awkward to use kilometers to express the large distances to planets, stars, or galaxies as it is to talk about the distance from New York City to San Francisco or Sydney to Perth in millimeters. Astronomers have therefore devised new units of measure.

When discussing distances across the solar system, astronomers use a unit of length called the astronomical unit (AU), which is the average distance between Earth and the Sun:

$$1 \text{ AU} \approx 1.5 \times 10^8 \text{ km} \approx 9.3 \times 10^7 \text{ mi}$$

Jupiter, for example, is an average of 5.2 times farther from the Sun than is Earth. Thus, the average distance between the Sun and Jupiter can be conveniently stated as 5.2 AU. This value can be converted into kilometers or miles using the previous relationship.

When talking about distances to the stars, astronomers choose between two different units of length. One is the **light-year (ly)**. A light-year is the *distance* that light travels in a year through a vacuum (that is, in the absence of air, glass, or other medium). Do keep in mind that the word *year* in this unit helps describe a separation between two objects rather than representing a unit of time.

$$1 \text{ ly} \approx 9.46 \times 10^{12} \text{ km} \approx 63,200 \text{ AU}$$

The spaces between the planets, stars, and galaxies are nearly ideal vacuums. One light-year is roughly equal to 6 trillion miles. Proxima Centauri, the closest star to Earth, other than the Sun, is just over 4.2 ly from us.

The second commonly used unit of length is the **parsec (pc)**, the distance at which two objects separated by 1 AU make an angle of 1 arcsec (denoted 1"). Imagine taking a journey far into space, beyond the orbits of the outer planets. Watching the solar system as you move away, the angle between the Sun and Earth becomes smaller and smaller. When they are side by side from your perspective, and you measure the angle between them as $1/3600^\circ = 1''$, you have reached a distance that astronomers call 1 parsec, as shown in the accompanying figure. The parsec turns out to be longer than the light-year, specifically,

$$1 \text{ pc} \approx 3.09 \times 10^{13} \text{ km} \approx 3.26 \text{ ly}$$

Thus, the distance to the nearest star can be stated as 1.3 pc as well as 4.2 ly. Whether one uses light-years or parsecs is a matter of personal preference.

For larger distances, *kilolight years (kly)*, *megalight years (Mly)*, *kiloparsecs (kpc)*, and *megaparsecs (Mpc)* are used. The prefixes *kilo* and *mega* simply mean “thousand” and “million,” respectively:

$$1 \text{ kly} = 10^3 \text{ ly}$$

$$1 \text{ Mly} = 10^6 \text{ ly}$$

$$1 \text{ kpc} = 10^3 \text{ pc}$$

$$1 \text{ Mpc} = 10^6 \text{ pc}$$

For example, the distance from Earth to the center of our Milky Way Galaxy is about 8.6 kpc, and the rich cluster of galaxies in the direction of the constellation Virgo is 20 Mpc away.

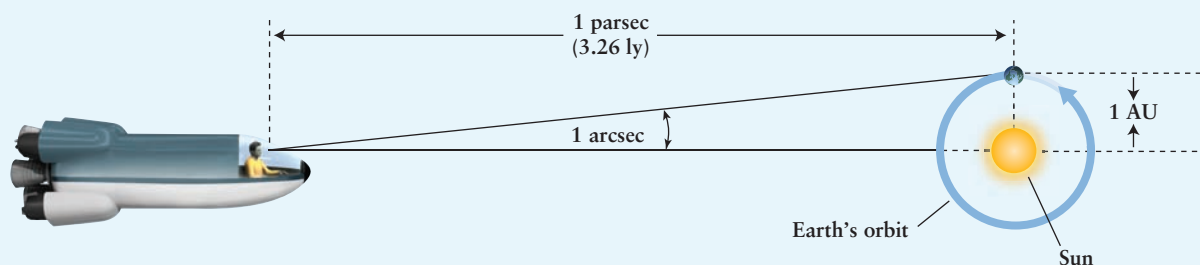


FIGURE D2-2 A Parsec The parsec, a unit of length commonly used by astronomers, is equal to 3.26 ly. The parsec

is defined as the distance at which 1 AU perpendicular to the observer's line of sight makes an angle of 1 arcsec.

When Newton derived Kepler's third law using the law of gravitation, discussed later in this chapter, he discovered that the mass of a planet affects the period of its orbit around the Sun. The **mass** of an object is a measure of the total number of particles of all different types that

it contains and is expressed in units of kilograms. For example, the mass of the Sun is 2×10^{30} kg, the mass of a hydrogen atom is 1.7×10^{-27} kg, and the mass of the author of this book is 83 kg. At rest, the Sun, a hydrogen atom, and I have these same masses regardless of where

we happen to be in the universe. It is important not to confuse the concept of mass with the concept of weight. Your **weight** is the force with which you push down on a scale due to the gravitational attraction of the world on which you stand.

However, the effect of the planet's mass on the period of its orbit is exceedingly small for all the planets in the solar system, which is why the equation for Kepler's third law, as shown in Table 2-3, gives such good results for the planets' orbits even though it does not take their masses into account. When calculating the motion of pairs of stars orbiting each other, the effects of the masses must be included, as described in **Discovery 9-3: Kepler's Third Law and Stellar Masses**.

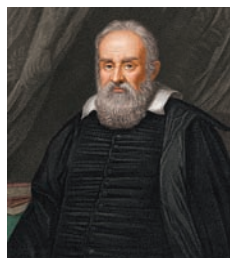
Kepler's three laws apply not only to the planets orbiting the Sun but also to any object orbiting another under the influence of their mutual gravitational attraction. Thus, Kepler's laws apply to moons orbiting planets, artificial satellites orbiting Earth, and even (with the above caveat) two stars revolving around each other.

2-6 Galileo's discoveries strongly supported a heliocentric cosmology

While Kepler was in central Europe working on the laws of planetary orbits, an Italian physicist was making dramatic observations in southern Europe. Galileo Galilei did not invent the telescope, but he was one of the first people to point the new device toward the sky and publish his observations. He saw things that no one had ever imagined—mountains on the Moon and spots on the Sun. He also discovered that like our Moon, Venus has phases from new to full and that the change in the apparent size of Venus as seen through his telescope was related to the planet's phase (**Figure 2-10**). Venus appears smallest at full phase and largest at new phase. These observations were a big chink in the geocentric cosmology's armor, as that model could not explain why Venus has a complete cycle of phases or changes size, while a heliocentric cosmology explains both. (Since Venus is always near the Sun in our sky, the geocentric model predicts that Venus should always have a crescent or quarter phase if it is closer to us than the Sun or a full or gibbous phase if it is farther away than the Sun—

see Figure 2-10.) Galileo's observations of a complete cycle of phases, therefore, supported the conclusion that Venus orbits the Sun, not Earth.

In 1610, Galileo also discovered four moons near Jupiter. Today, in honor of their discoverer, these are called the **Galilean moons** (or **satellites**, another term for moons). Galileo concluded that the moons



Galileo Galilei
(1564–1642) (Stock
Montage, Inc./Alamy)

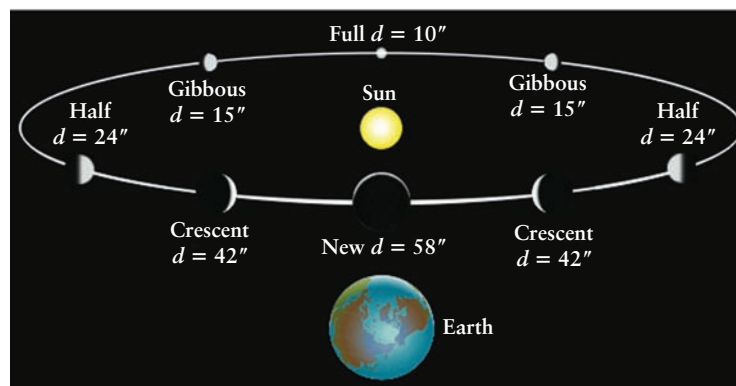


FIGURE 2-10 The Changing Appearance of Venus This figure shows how the appearance (phase) of Venus changes as it moves along its orbit. The number below each view is the angular diameter (d) of the planet as seen from Earth, in arcseconds. The " indicates arcseconds, as introduced in Section 1-5 and again in **Discovery 2-2: Units of Astronomical Distance**. Note that the phases correlate with the planet's angular size and its angular distance from the Sun, both as seen from Earth. These observations clearly support the idea that Venus orbits the Sun.

orbit Jupiter because he saw them move in straight lines from one side of the planet to the other. (He did not see them move in elliptical orbits because from Earth we see their orbits from edge-on.) Confirming observations were made in 1620 (**Figure 2-11**). These observations all provided further evidence that Earth is not at the center of the universe. Like Earth in orbit around the Sun, Jupiter's four moons obey Kepler's third law: The square of a moon's orbital period around Jupiter is directly proportional to the cube of its average distance from the planet.

Galileo's telescopic observations constituted the first fundamentally new astronomical data since humans began recording what they saw in the sky. In contradiction to then-prevailing opinions, these discoveries strongly supported a heliocentric view of the universe. Because Galileo's ideas could not be reconciled with certain passages in the Bible or with the writings of Aristotle and Plato, the Roman Catholic Church condemned him, and he was forced to spend his later years under house arrest "for vehement suspicion of heresy."

Insight Into Science

Theories and Explanations Scientific theories (or laws) based on observations can be useful for making predictions even if the reasons that these theories work are unknown. The explanation for Kepler's laws came decades after Kepler deduced them, when Newton derived them in 1665 with his mathematical expression for gravitation, the force that holds the planets in their orbits.



a

In 1992, Pope John Paul II stated that the Church erred in this condemnation.

A major stumbling block prevented seventeenth-century thinkers from accepting Kepler's laws and Galileo's conclusions about the heliocentric cosmology. Once anything on Earth is put in motion, it quickly comes to rest. So why don't the planets orbiting the Sun stop, too?

The scientific method clarified most of the issues surrounding planetary orbits, leading to the equations and laws developed by the brilliant and eccentric scientist Isaac Newton (for example, he believed in alchemy). He was born on Christmas Day in 1642, less than a year after Galileo died. In the decades that followed, Newton revolutionized science more profoundly than any person before him and, in doing so, he found physical and mathematical evidence in support of the heliocentric cosmology.

2-7 Newton formulated three laws that describe fundamental properties of physical reality

Until the mid-seventeenth century, virtually all mathematical astronomy was done empirically. That is, astronomers from Ptolemy to Kepler created equations directly from data and observations.

Isaac Newton introduced a new approach. He began with three physical assumptions, now called **Newton's laws of motion**, which led to equations that have since been tested and shown to be correct in many everyday

FIGURE 2-11 Jupiter and Its Largest Moons In 1610, Galileo discovered four “stars” that move back and forth across Jupiter. He concluded that they are four moons that orbit Jupiter just as our Moon orbits Earth. **(a)** Observations made by Jesuits in 1620 of Jupiter and its four visible moons. **(b)** Photograph of the four Galilean satellites alongside an overexposed image of Jupiter. Each satellite would be bright enough to be seen with the unaided eye were it not overwhelmed by the glare of Jupiter. (Rev. Ronald Royer/Science Source)



b

RIVUX6

situations. He also found a formula for the force of **gravity** (or *gravitation*), the attraction between all objects due to their masses. Putting the assumptions into mathematical form and combining them with the equation for gravity, Newton was able to derive Kepler's three laws and use them to predict the orbits of bodies such as comets and other objects in the solar system. Newton also was able to use these same equations to predict the motions of bodies on and near Earth, such as the path of a projectile or the speed of a falling object.

Newton's First Law—The Law of Inertia: *Inertia is the property of matter that keeps an object at rest or moving in a straight line at a constant speed unless acted upon by a net external force.*

If all of the external forces acting on an object do not cancel each other out, then there is a net external force acting on the object. Equivalently, we say that there is an unbalanced external force. For example, if you put a soccer ball between your hands and press on it so that it does not move, your hands represent a balanced pair of forces acting on the ball. In that case, you are exerting no net external force on the ball. Conversely, when your foot hits a soccer ball and the ball sails away, your foot has exerted a net external force on the ball.

At first, this law might seem to conflict with your everyday experience. For example, if you shove a chair, it does not move at a constant speed forever but comes to rest after sliding only a short distance. From Newton's



Isaac Newton (1642–1727) (Lebrecht Music and Arts Photo Library/Alamy)

viewpoint, however, a “net external force” does indeed act on the moving chair—namely, friction between the chair’s legs and the floor. Without friction, the chair would continue in a straight path at a constant speed. A net external force changes the motion of an object.

Newton’s first law tells us why the planets keep moving in orbit around the Sun. First, they do not come to rest because there is virtually no air in space and hence

no force from, for example, air friction opposing their motion. Second, they do not move in straight lines because there is an outside force acting on the planets to continually change their directions and keep them in orbit. As we shall see, that force is the Sun’s gravity.

Newton’s second law describes quantitatively how a force changes the motion of an object. To better appreciate the concepts of force and motion, we must first understand two related quantities: velocity and acceleration.

Imagine an object motionless in space. Push on it and it begins to move. At any moment, you can describe the object’s motion by specifying both its speed and direction. Speed and direction of motion together constitute an object’s **velocity**. If you continue to push on the object, its speed will increase—it will accelerate.

Acceleration is the rate at which velocity changes with time. Because velocity involves both speed and direction, a slowing down, a speeding up, and a change in direction are all forms of acceleration.

Suppose, for example, an object revolved around the Sun in a perfectly circular orbit. As this object moved along its orbit, its speed would remain constant, but its direction of motion would be continuously changing. This body would have acceleration that involved only a change of direction. In general:

Newton’s Second Law—The Force Law: *The acceleration of an object is directly proportional to the net force acting on it and is inversely proportional to its mass.*

In other words, the harder you push on something that can move, the faster it will accelerate. Also, an object of greater mass accelerates more slowly when acted on by a force than does an object of lesser mass acted on by the same force. That is why you can accelerate a child’s wagon faster than you can accelerate a car by pushing on them equally hard.

Newton’s second law can be succinctly stated as an equation. If a **force** acts on an object, the object will experience an acceleration such that

$$\text{Force} = \text{mass} \times \text{acceleration}$$

5 Force is usually expressed in pounds or newtons. For example, the force with which I am pressing down on the ground is 814 newtons (183 lb). But I weigh 814 newtons only on Earth. I would weigh 136 newtons (30.5 lb) on the Moon, which has less mass and so pulls me down with less gravitational force. Orbiting in the International Space Station, my weight (measured by standing on a scale in the space station) would be 0, but my mass would be the same as when I am on Earth. Whenever we describe the properties of planets, stars, or galaxies, we speak of their masses, never of their weights.

Newton’s final assumption, called *Newton’s third law*, is the law of action and reaction.

Newton’s Third Law—The Law of Action and Reaction: *Whenever one object exerts a force on a second object, the second object exerts an equal and opposite force on the first object.*

For example, I weigh 183 lb on Earth, and so I press down on the floor with a force of 183 lb. Newton’s third law says that the floor is also pushing up against me with an equal force of 183 lb. (If it were less, I would fall through the floor, and if it were more, I would be lifted upward.) Even two objects that are not touching exert equal and opposite gravitational forces on each other. The Sun is exerting a force on each planet to keep it in orbit; hence, each planet exerts an equal and opposite force on the Sun. As each planet accelerates toward the Sun, the Sun in turn accelerates toward each planet.

Because the Sun is pulling on the planets, why don’t they fall onto it? Conservation of angular momentum provides the answer. Angular momentum is a measure of how much energy is stored in an object due to its rotation and revolution. The details of momentum are presented in Appendix P: Energy and Momentum. As the orbiting planets fall toward the Sun, their angular momentum provides them with motion perpendicular to that infall, meaning that the planets continually fall toward the Sun, but they continually miss it. Because their angular momentum is conserved, planets neither spiral into the Sun nor fly away from it. Angular momentum remains constant unless acted on by an external torque (also defined in Appendix P).

Angular momentum depends on three things: how fast an object rotates or revolves, how much mass it has, and how spread out that mass is. Consider, for example, a twirling ice

Focus Question 2-7

Sitting in a moving car, how can you experimentally verify that your body has inertia?

Focus Question 2-8

How did people deduce that there is no air (and, hence, no air friction to slow planets down) in space before airplanes or even people-carrying balloons were invented?

Focus Question 2-9

If you are on a freely spinning merry-go-round, what will happen to it as you move toward the center?

FIGURE 2-12 Conservation of Angular Momentum As skater Sarah Hughes brings her arms and outstretched leg in, she must spin faster to conserve her angular momentum. You can see that motion in her skirt and hair. (Doug Pensinger/Getty Images Sport/Getty Images)



skater. She rotates with a constant mass, practically free of outside forces. Because her angular momentum is therefore conserved, she can change how fast she is spinning by changing how spread out her mass is. According to conservation of angular momentum, as the spread of mass decreases, the rotation rate must increase. When she wishes to rotate more rapidly, she decreases the spread of her mass distribution by pulling her arms and outstretched leg in closer to her body (**Figure 2-12**). In astronomy, we encounter many instances of the same law, as giant objects, such as stars, contract.

We have now reconstructed the central relationships between matter and motion. Scientific explanation of the heliocentric cosmology still requires a force to hold the planets in orbit around the Sun and the moons in orbit around the planets. Newton identified that, too.

gravitational force from every object acts to pull every other object directly toward it.

Newton succeeded in formulating a mathematical model that describes the behavior of the gravitational force that keeps the planets in their orbits (presented in Appendix R: Gravitational Force).

Newton’s Law of Universal Gravitation: *Two objects attract each other with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.*

The inverse square part of this law means that gravitational force decreases with distance. **Table 2-4** gives some examples of the inverse square law. Despite its weakening, the force of gravity from each object extends throughout

2-8 Newton’s description of gravity accounts for Kepler’s laws

Isaac Newton did not invent the idea of gravity. An observant seventeenth-century person would understand that some force pulls things down to the ground. It was Newton, however, who gave us a quantitative description of the action of gravity, or *gravitation*, as it is more properly called. Using his first two laws, Newton showed mathematically that the force acting on each of the planets is directed toward the Sun. He expanded this result to the idea that the nature of the force pulling a falling apple straight down to the ground is the same as the nature of the force on the planets from the Sun. More generally, the

TABLE 2-4	
EXAMPLES OF THE INVERSE SQUARE LAW FOR GRAVITY. Inverse square means the force of gravity decreases as 1/(distance between two objects) ² .	
Distance between two objects	The gravitational force between two objects compared to when they are a distance of 1 apart
1	1
2	1/4
3	1/9
4	1/16

the universe. Also, an object with twice the mass of another object exerts twice the gravitational force as the less massive object.

Using his law of gravity along with his three laws stated earlier, Newton found that he could mathematically explain Kepler's three laws. For example, whereas Kepler discovered by trial and error that the period of orbit, P , and average distance between the Sun and planet, a , are related by $P^2 = a^3$, Newton mathematically derived this equation (corrected by including a tiny contribution due to the mass of the planet, as mentioned earlier). Bodies in elliptical orbits are bound by the force of gravity to remain in orbit.

6 It seems plausible that astronauts floating in the International Space Station do not feel any force of gravity from Earth, but they do. Orbiting 330 km (approximately 200 mi) above Earth's surface, they feel 90% as much gravitational force from the planet as we do standing on it. They are weightless, however, because as they fall earthward, their angular momentum carries them around the planet at just the right rate to continually miss it. The same effect occurs for all objects in orbit, as shown in the figure that opens this chapter.

Newton also discovered that some objects orbiting the Sun can follow nonelliptical paths. His equations led him to conclude that orbits can also be **parabolas** or **hyperbolas** (Figure 2-13). In both cases, such bodies would make only one pass close to the Sun and then travel out of the solar system, never to return. To date, all of the objects observed in the solar system began their existence in elliptical orbits, but some comets (small bodies of rock and ice) have received enough energy from being pulled by planets or from expelling jets of gas to develop parabolic or hyperbolic orbits.

Using the equations Newton derived, the orbits of the planets and their satellites could be calculated with unprecedented precision. Using his laws, mathematicians showed that Earth's axis of rotation must precess because of the gravitational pull of the Moon and the Sun on Earth's equatorial bulge (recall Figure 1-19). In the spirit of the scientific method, Newton's laws and mathematical techniques were used to predict new phenomena. For example, Edmond Halley was intrigued by historical records of a comet that was sighted about every 76 years. Using his friend Newton's methods, Halley worked out the details of the comet's orbit and predicted its return in 1758. It was first sighted on Christmas night of that year, and to this day the comet bears Halley's name (Figure 2-14).

Perhaps the most dramatic early use of the scientific method with Newton's ideas was its role in the discovery of the eighth planet in our solar system. The seventh planet, Uranus, had been discovered by William Hersch in 1781 during a systematic telescopic survey of the sky. Fifty years later, however, it was clear that Uranus was

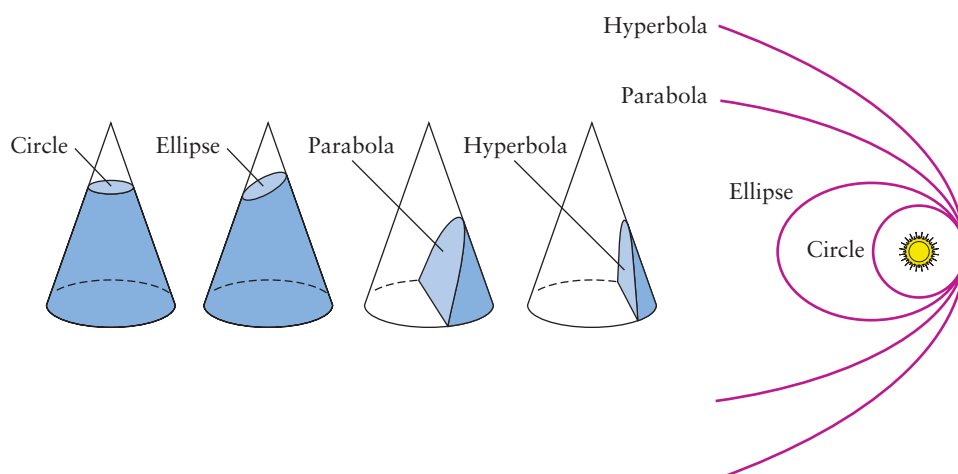


FIGURE 2-13 Conic Sections A conic section is any one of a family of curves obtained by slicing a cone with a plane, as shown. The orbit of one body around another can be an ellipse, a parabola, or a hyperbola. Circular orbits are possible because a circle is just an ellipse for which both foci are at the same point.

not following the orbit predicted by Newton's laws. Two mathematicians, John Couch Adams in England and Urbain-Jean-Joseph Leverrier in France, independently calculated that the deviations of Uranus from its predicted orbit could be explained by the gravitational pull of a then unknown, more distant planet. Each man predicted that the planet would be found at a certain location in the constellation of Aquarius in September 1846. A telescopic search on September 23, 1846, by German astronomer Johann Galle, revealed Neptune less than 1° from its calculated position. Although sighted with a telescope, Neptune was really discovered with pencil and paper.



FIGURE 2-14 Halley's Comet Halley's Comet orbits the Sun with an average period of about 76 years. During the twentieth century, the comet passed near the Sun twice—once in 1910 and again, as shown here, in 1986. The comet will pass close to the Sun again in 2061. During its last visit, the comet spread more than 5° across the sky, or 10 times the diameter of the Moon. (W. Liller/Large Scale Phenomena Network/NSSDC/NASA)

Insight Into Science

Quantify Predictions Mathematics provides a language that enables science to make quantitative predictions that can be checked by anyone. For example, in this chapter, we have seen how Kepler's third law and Newton's universal law of gravitation correctly predict the motion of objects under the influence of the Sun's gravitational attraction.

It is a testament to Newton's genius that his three laws were precisely the basic ideas needed to understand so much about the natural world. Newton's process of deriving Kepler's laws and the universal law of gravitation helped secure the scientific method as an invaluable tool in our process of understanding the universe. **Figure 2-15** shows some of the effects of gravity at the scales of planets, stars, and galaxies.

SUMMARY OF KEY IDEAS

Science: Key to Comprehending the Cosmos

- The ancient Greeks laid the groundwork for progress in science by stating that the universe is comprehensible.
- The scientific method is a procedure for formulating theories that correctly predict how the universe behaves.
- A scientific theory must be testable, that is, capable of being disproved.
- Theories are tested and verified by observation or experimentation and result in a process that often leads to their refinement or replacement and to the progress of science.
- Observations of the cosmos have led astronomers to discover some fundamental physical laws of the universe.

Changing Our Earth-Centered View of the Universe

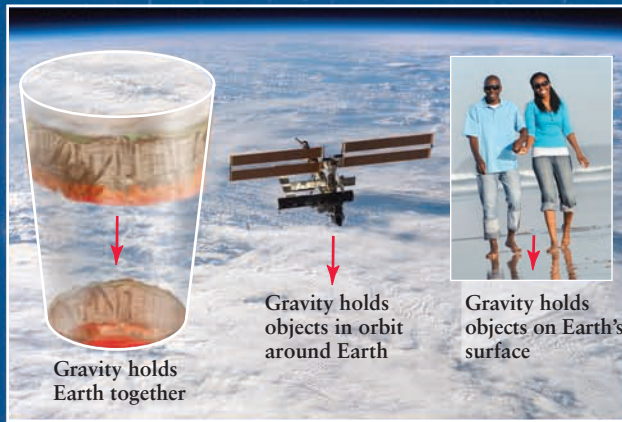
- Common sense (for example, Earth doesn't appear to be moving) led early natural philosophers to devise a geocentric cosmology, which placed Earth at the center of the universe.
- Kepler modified Copernicus's heliocentric (Sun-centered) theory by showing that orbits are elliptical, thereby creating a simplified explanation of planetary motions compared to the geocentric theory.
- The heliocentric cosmology refers to motion of planets and smaller debris orbiting the Sun. Other stars do not orbit the Sun.
- The sidereal orbital period of a planet is measured with respect to the stars and determines the length of the planet's year. A planet's synodic period is measured with respect to the Sun as seen from the moving Earth (for example, from one opposition to the next).

Kepler's and Newton's Laws

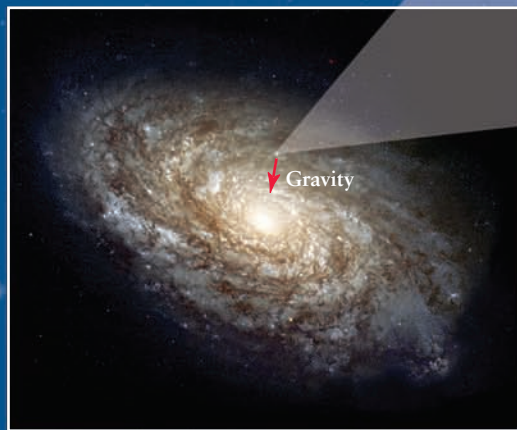
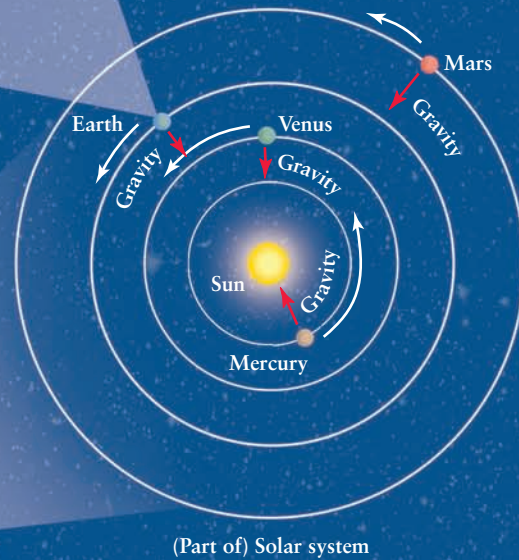
- Ellipses describe the paths of the planets around the Sun much more accurately than do the circles used in previous theories. Kepler's three laws give important details about elliptical orbits.
- The invention of the telescope led Galileo to new discoveries, such as the phases of Venus and the moons of Jupiter, which supported a heliocentric view of the universe.
- Newton based his explanation of the universe on three assumptions, now called Newton's laws of motion. These laws and his law of universal gravitation can be used to deduce Kepler's laws and to describe most planetary motions with extreme accuracy.
- The mass of an object is a measure of the amount of matter in it; weight is a measure of the force with which the gravity of a world pulls on an object's mass when the two objects are at rest with respect to each other (or, equivalently, how much the object pushes down on a scale).
- The path of one astronomical object around another, such as that of a comet around the Sun, is an ellipse, a parabola, or a hyperbola. Ellipses are bound orbits, while objects with parabolic and hyperbolic orbits fly away, never to return.

WHAT DID YOU THINK?

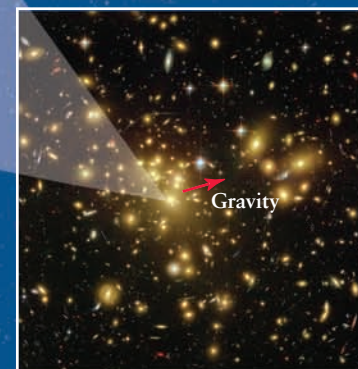
- 1** *What makes a theory scientific?* A theory is an idea or set of ideas proposed to explain something about the natural world. A theory is scientific if it makes predictions that can be objectively tested and potentially disproved.
- 2** *What is the shape of Earth's orbit around the Sun?* All planets have elliptical orbits around the Sun.
- 3** *Do the planets orbit the Sun at constant speeds?* No. The closer a planet is to the Sun in its elliptical orbit, the faster it is moving. The planet moves fastest at perihelion and slowest at aphelion.
- 4** *Do all of the planets orbit the Sun at the same speed?* No. A planet's speed depends on its average distance from the Sun. The closest planet moves fastest, while the most distant planet moves slowest.
- 5** *How does an object's mass differ when measured on Earth and on the Moon?* Assuming the object doesn't shed or collect pieces, its mass is the same on Earth and on the Moon. Its weight, however, is less on the Moon.
- 6** *Do astronauts orbiting Earth feel the force of gravity from our planet?* Yes. They are continually pulled earthward by gravity, but they continually miss it because of their motion around it.



Earth



Galaxy



Cluster of galaxies

FIGURE 2-15 Gravity Works at All Scales This figure shows a few of the effects of gravity here on Earth, in the solar system, in our Milky Way Galaxy, and beyond. The arrow in the cluster of galaxies shows the direction of the force of gravity from one

cluster (bright group of galaxies on the right) on another cluster of galaxies. (Space station: NASA; couple holding hands: Warren Goldswain/Shutterstock; galaxy: NASA; galaxy cluster: ESA, NASA, J.-P. Kneib [Caltech/Observatoire Midi-Pyrénées] and R. Ellis [Caltech])

Review Questions

- Who wrote down the equation for the law of gravitation?
 - Copernicus
 - Tycho
 - Newton
 - Galileo
 - Kepler
- Which of the following most accurately describes the shape of Earth's orbit around the Sun?
 - circle
 - ellipse
 - parabola
 - hyperbola
 - square
- Of the following planets, which takes the longest time to orbit the Sun?
 - Earth
 - Uranus
 - Mercury
 - Jupiter
 - Venus
- What is a Sun-centered model of the solar system called?
- How long does it take Earth to complete a sidereal orbit of the Sun?
- How did Copernicus explain the retrograde motions of the planets?
- Which planets can never be seen at opposition? Which planets never pass through inferior conjunction?
- At what configuration (superior conjunction, greatest eastern elongation, etc.) would it be best to observe Mercury or Venus with an Earth-based telescope? At what configuration would it be best to observe Mars, Jupiter, or Saturn? Explain your answers.
- What are the synodic and sidereal periods of a planet?
- What are Kepler's three laws? Why are they important?
- In what ways did the astronomical observations of Galileo support a heliocentric cosmology?
- How did Newton's approach to understanding planetary motions differ from that of his predecessors?
- What is the difference between mass and weight?
- Why was the discovery of Neptune a major confirmation of Newton's universal law of gravitation?
- Why does an astronaut have to exert a force on a weightless object to move it?

Got It?

- A comet coming inward from the Kuiper belt, a region of Sun-orbiting debris out beyond the orbit of Neptune, experiences a gravitational force from the Sun. Does the presence of the planets affect the comet's orbit? Explain your reasoning.
- How would the weight of an astronaut on the Moon compare to her weight on Earth?
- How would the mass of an astronaut on the Moon compare to his mass on Earth?
- An astronomer observes a new comet and calculates that it will exit the solar system and not return. Which of the following best describes the path of the comet?
 - a nearly straight line
 - a circle
 - an ellipse
 - a hyperbola
 - some other shape

Advanced Questions

- From the definition $KE = \frac{1}{2}mv^2$, derive the equation $KE = p^2/2m$, as discussed in Appendix P.
- Is it possible for an object in the solar system to have a synodic period of exactly 1 year? Explain your answer.
- Describe why there is a systematic decrease in the synodic periods of the planets from Mars outward, as shown in Table 2-1.
- Make diagrams of Jupiter's phases as seen from Earth and as seen from Saturn.
- In what direction (left or right, eastward or westward) across the celestial sphere do the planets normally appear to move as seen from Australia? In what direction is retrograde motion as seen from there?
- The dictionary defines astrology as "the study that assumes and attempts to interpret the influence of the heavenly bodies on human affairs." Based on what you know about scientific theory, is astrology a science? Why or why not? Feel free to further explore astrology, if you wish, before answering this question.

Discussion Questions

- Which planet would you expect to exhibit the greatest variation in apparent brightness as seen from Earth? Explain your answer.
- Use two thumbtacks (or pieces of tape), a loop of string, and a pencil to draw several ellipses. Describe how the shapes of the ellipses vary as you change the distance between the thumbtacks.



RIVUUXG Three of the four buildings containing the Very Large Telescope optical telescopes at the Paranal Observatory, Atacama, Chile. (Babak Tafreshi / Science Source)

Light and Telescopes

WHAT DO YOU THINK?

- 1 What is light?
- 2 Which type of electromagnetic radiation is most dangerous to life?
- 3 What is the main purpose of a telescope?
- 4 Why do all research telescopes use mirrors, rather than lenses, to collect light?
- 5 Why do stars twinkle?
- 6 Which is hotter, a "red-hot" or a "blue-hot" object?
- 7 What color does the Sun emit most brightly?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

Our understanding of the universe comes primarily from visible light and other forms of radiation collected by telescopes. We begin this chapter by exploring the nature of this radiation. Then we study telescopes. Finally, we examine how light and other radiations are created and emitted.

In this chapter you will discover

- the connection between visible light, radio waves, X-rays, and other types of electromagnetic radiation
- the debate in past centuries over what light is and how Einstein resolved this question
- how telescopes collect and focus light
- the limitations of telescopes, especially those that use lenses to collect light
- why different types of telescopes are used for different types of research
- how astronomers use the entire spectrum of electromagnetic radiation to observe the stars and other astronomical objects and events
- what the new generations of land-based and space-based high-technology telescopes being developed can do
- the origins of electromagnetic radiation
- the structure of atoms
- that stars with different surface temperatures emit different intensities of electromagnetic radiation
- that astronomers can determine the chemical compositions of stars and interstellar clouds by studying the wavelengths of electromagnetic radiation that they absorb or emit
- how to tell whether an object in space is moving toward or away from Earth

THE NATURE OF LIGHT

So far in this text we have used the word *light* in its everyday sense—the stuff to which our eyes are sensitive. This is more properly called *visible light*, and it is a form of **electromagnetic radiation**. Perhaps contrary to one's intuition, this radiation is composed of particles, called *photons*, that have properties of waves. Detecting electromagnetic radiation with telescopes is the essence of observational astronomy. Although human perception of objects here on Earth and in space comes primarily from the visible light that our eyes detect, visible light is only a tiny fraction of all the electromagnetic radiation emitted by objects in the universe. The rest of this radiation, invisible to our eyes, is detected by high-tech sensors attached to specially designed telescopes. In this chapter, we examine telescopes for all types of electromagnetic radiation. To understand how telescopes work, we begin by exploring the properties of the electromagnetic radiation that they collect.

3-1 Newton discovered that white is not a fundamental color and proposed that light is composed of particles

From the time of Aristotle (in the fourth century B.C.E.) until the late seventeenth century C.E., most people believed that white is the fundamental color of light. The colors of the rainbow (or, equivalently, the colors created by light passing through a prism) were believed to be added or created somehow as white light went from one medium through another. Isaac Newton performed experiments during the late 1600s that disproved these beliefs. He started by passing a beam of sunlight

through a glass prism, which spread the light out into the colors of the rainbow (**Figure 3-1a**). The change in direction as light travels from one medium into another is called **refraction**, and the resulting spread of colors (complete or with colors missing) is called a **spectrum** (plural, **spectra**).

Then Newton selected a single color and sent it through a second prism (**Figure 3-1b**). The light that emerged from the first prism was refracted by the second prism, but *it remained the same color*. The fact that individual colors of refracted light were unchanged by the second prism led Newton to conclude that the colors of a full spectrum (shades of red, orange, yellow, green, blue, and violet; indigo is not considered a separate color in astronomy) were in fact properties of the light itself, and that white light is a mixture of colors. To prove this last point, he recombined all the spectrum colors, thereby recreating white light. Thus, different colors of the spectrum are different entities. But what, Newton wondered, is the nature of light that it could break apart into distinct colors and then be completely reconstituted?

Back in the mid-1600s, the Dutch scientist Christiaan Huygens proposed that light travels in the form of waves. Newton, on the other hand, performed many experiments in optics that convinced him that light is composed of tiny particles of energy. It turns out that both ideas were right.

In 1801, the English physicist Thomas Young demonstrated that light is indeed composed of waves. Young sent light of a single color through two parallel slits (**Figure 3-2a**). He reasoned that if the light were waves, then these waves would behave the way waves

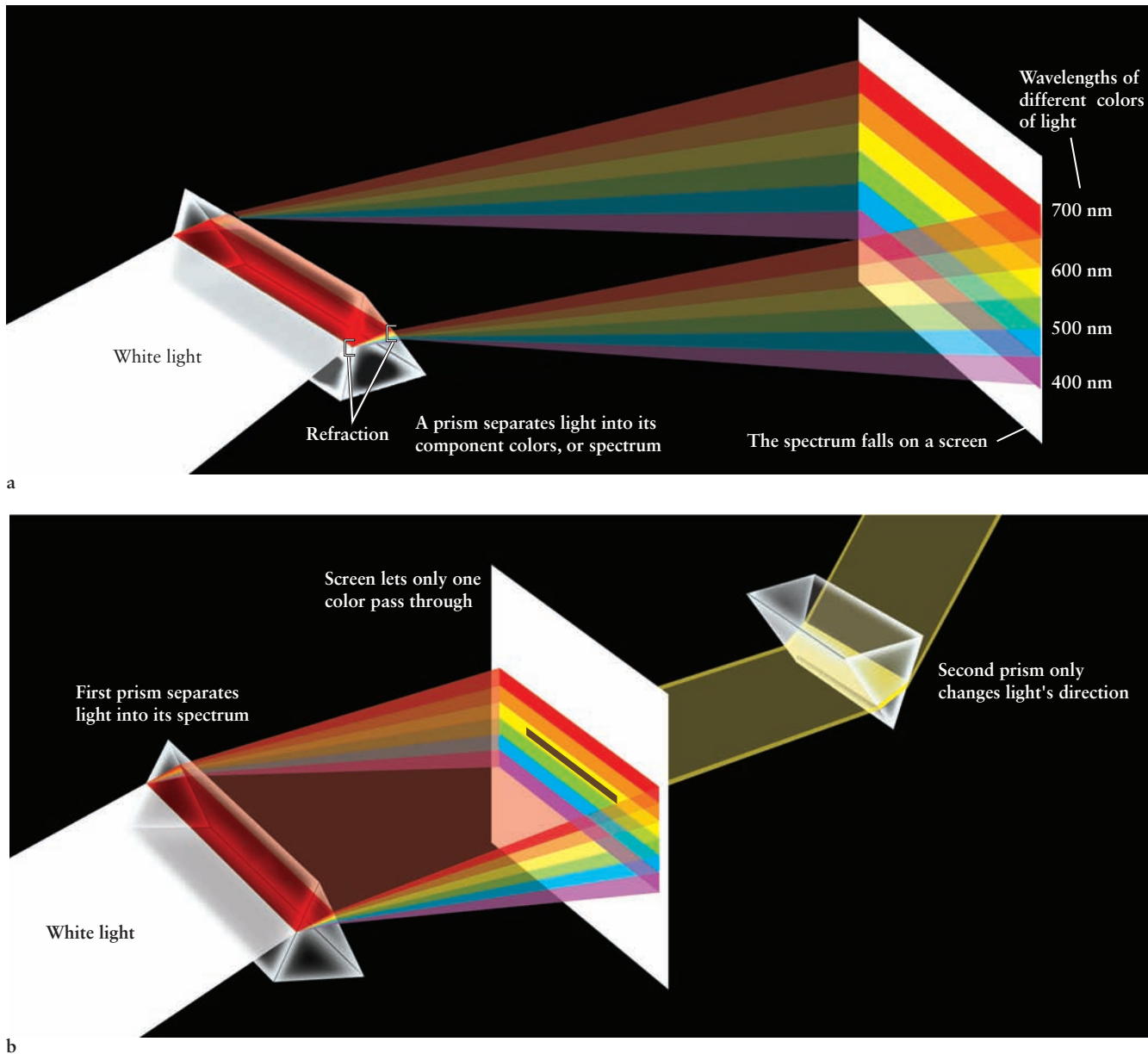


FIGURE 3-1 Prisms and a Spectrum (a) When a beam of white light passes through a glass prism, the light is separated or refracted into a rainbow-colored band called a spectrum. The numbers on the right side of the spectrum indicate wavelengths

in nanometers ($1 \text{ nm} = 10^{-9} \text{ m}$). (b) This drawing of Newton's experiment illustrates that glass does not add to the color of light, only changes its direction. Because color is not added, this experiment shows that color is an intrinsic property of light.

on the surface of water behave, flowing through similar gaps (Figure 3-2b). In particular, he theorized that the light waves from each slit would interact with light waves from the other. For example, when two light waves meet, with one wave going up and the other going down, they would interfere with each other and partially or totally cancel each other out, leaving dark regions on the screen. When two waves that are both moving up or both moving down meet, they would reinforce each other and create bright regions. As you can see in Figure 3-2, this is precisely what happened. (If

light were just random particles going through the slits, they would not interfere with each other in this way, and the pattern on the screen would just yield two bright regions, one behind each slit.) The analogy with water waves ends here. For example, light waves do not have to travel through a medium, unlike water waves, which travel through the liquid medium of water.

Further insight into the wave character of light came from calculations by the Scottish physicist James Clerk Maxwell in the 1860s. Maxwell unified the descriptions of the basic properties of electricity and magnetism into

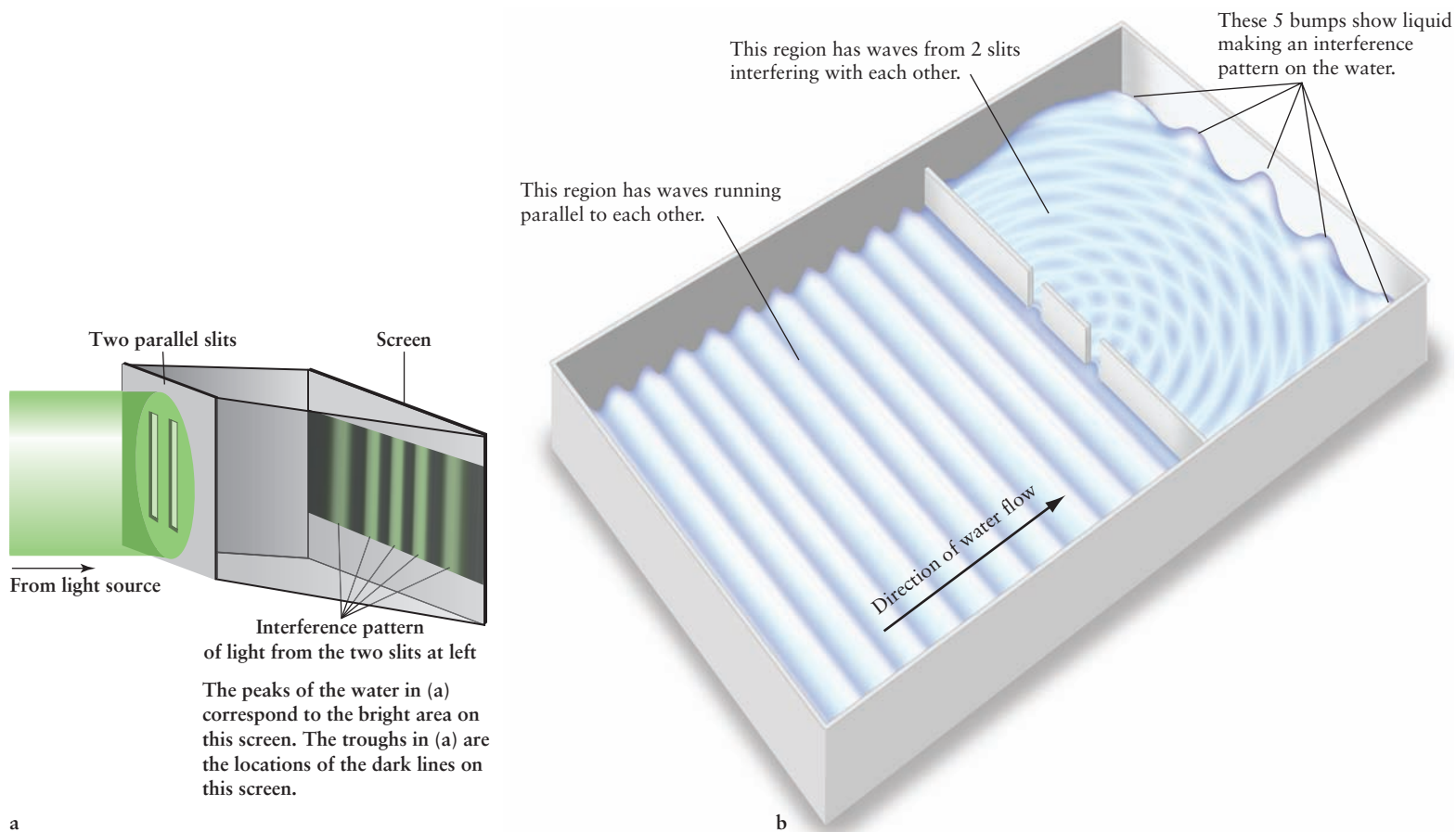


FIGURE 3-2 Wave Travel (a) Electromagnetic radiation travels as waves. Thomas Young's interference experiment shows that light of a single color passing through a barrier with two slits behaves as waves that create alternating light and dark patterns on a screen. **(b)** Water waves passing through two slits in a ripple tank create interference patterns. The wa-

ter waves interfere with each other, creating constructive interference (crests) and destructive interference (troughs) throughout the right side of the tank and on the far right wall. (Original research by University of Colorado, Center for Integrated Plasma Studies, Boulder, CO)

four equations. By combining these equations, he demonstrated that electric and magnetic effects should travel through space together in the form of coupled waves (**Figure 3-3**) that have equal amplitudes. Maxwell's suggestion that some of these waves, now called *electromagnetic radiation*, are observed as visible light was soon confirmed by a variety of experiments. Despite the name electromagnetic radiation, neither visible light nor any other type of electromagnetic radiation is electrically charged.

Newton showed that sunlight is composed of all the colors of the rainbow. Young, Maxwell, and others showed that light travels as waves. What makes the colors of the rainbow distinct from each other? The answer is surprisingly simple: Different colors are waves with different wavelengths. A **wavelength**, usually designated by λ , the lowercase Greek letter lambda, is the distance between two successive wave crests (see Figure 3-3).

The wavelengths of all colors are extremely small, less than a thousandth of a millimeter. To express these tiny distances conveniently, scientists use a unit of length called the *nanometer* (nm), where $1 \text{ nm} = 10^{-9} \text{ m}$. Another unit you might encounter when talking to astronomers is the *angstrom* ($\text{\AA}$), where $1 \text{ \AA} = 0.1 \text{ nm} = 10^{-10} \text{ m}$. Experiments demonstrate that visible light has wavelengths ranging from about 400 nm for the shortest wavelength of violet light to about 700 nm for the longest wavelength of red light. Intermediate colors of the rainbow fall between these wavelengths (see Figure 3-1a). The complete spectrum of colors from the longest wavelength to the shortest is red, orange, yellow, green, blue, and violet. Referring to Figure 3-1, you can also see that the amount of refraction that different colors undergo depends on their wavelengths: *The shorter the wavelength, the more the light is refracted.*

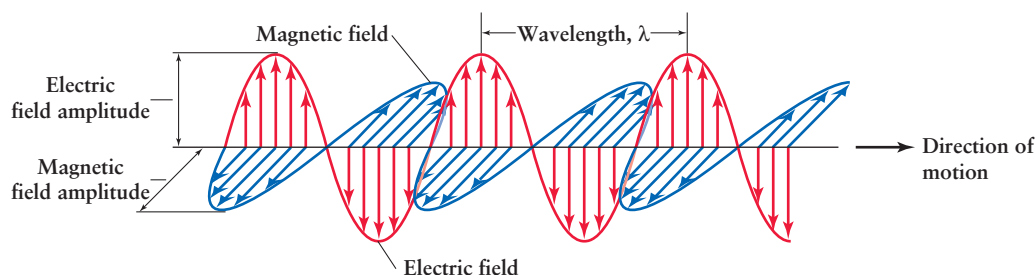


FIGURE 3-3 Electromagnetic Radiation All forms of electromagnetic radiation (radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays) consist of combined electric and magnetic fields oscillating perpendicular to each other and to the direction in which

they move. In empty space this radiation travels at a speed of 3×10^5 km/s. These fields are the mathematical description of the electric and magnetic effects. The distance between two successive crests, denoted by λ , is called the wavelength of the light.

3-2 Light travels at a finite but incredibly fast speed

The fact that we see lightning before we hear the accompanying thunderclap tells us that light travels faster than sound. But does that mean that light travels instantaneously from one place to another, or does it move with a measurable speed?

Light speed The first evidence for the finite speed of light came in 1675, when Ole Rømer, a Danish astronomer, carefully timed eclipses of Jupiter's moons (**Figure 3-4**). Rømer discovered that the moment at which a moon enters Jupiter's shadow depends on the distance between Earth and Jupiter. When Jupiter is in opposition—that is, when Jupiter and Earth are on the same side of the Sun (see **Figure 2-4**)—the Earth–Jupiter distance is relatively short compared to when Jupiter is near conjunction. At opposition, Rømer found that eclipses occur slightly earlier than predicted by Kepler's laws, and they occur slightly later than predicted when Jupiter is near conjunction.

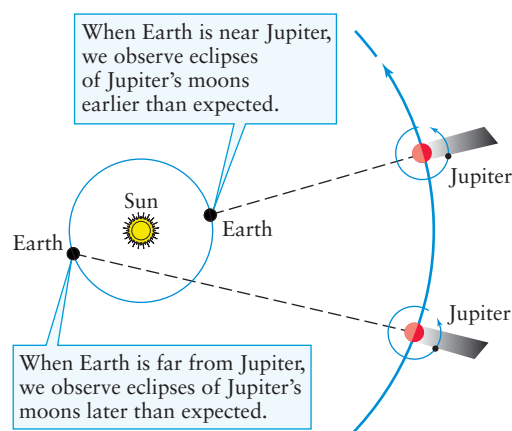


FIGURE 3-4 Evidence That Light Travels at a Finite Speed The times of the eclipses of Jupiter's moons as seen from Earth depend on the relative positions of Jupiter, Earth, and the Sun. Rømer correctly attributed the variations in these times to the variations in the time that it takes light from these events to reach Earth.

Rømer correctly concluded that light travels at a finite speed, and so it takes more time to travel longer distances across space. The greater the distance to Jupiter, the longer the image of an eclipse takes to reach our eyes. From his timing measurements, Rømer concluded that it takes 162 minutes for visible light to traverse the diameter of Earth's orbit (2 AU). Incidentally, Rømer's interpretation of the data requires a heliocentric cosmology—that both Earth and Jupiter orbit the Sun.

Focus Question 3-1

What color is refracted least?

Rømer's subsequent calculation of the speed of light was off by 25% because the value for the astronomical unit (the average distance from Earth to the Sun) that existed at that time was highly inaccurate. Nevertheless, he proved his main point—light travels at a finite speed. The first accurate laboratory measurements of the speed of visible light were performed in the mid-1800s.

Maxwell's equations also reveal that light of all wavelengths travels at the same speed in a vacuum (a region that contains no matter), and, despite a few atoms per cubic meter, the space between planets and stars is a very good vacuum. The constant speed of light in a vacuum, usually designated by the letter c , has been measured to be 299,792.458 km/s, which we generally round to

$$c = 3.0 \times 10^5 \text{ km/s} = 1.86 \times 10^5 \text{ mi/s}$$

(Standard abbreviations for units of speed, such as *km/s* for kilometers per second and *mi/s* for miles per second, will be used throughout the rest of this book.) Light that travels through air, water, glass, or any other substance always moves more slowly than it does in a vacuum.

The value c is a fundamental property of the universe. The speed of light appears in equations that describe, among other things, atoms, gravity, electricity, magnetism, distance, and time. Light has the extraordinary property that its speed is the same as seen by all observers. For example, if you were traveling in space away from the Earth at 99% of the speed of light, $0.99c$, you would measure the speed of light moving toward

you from all directions to be the same as the speed you would measure for that light if you were on Earth, namely, c . Likewise, you would measure the speed of light moving away from your spaceship in all directions to be the same as the speed of light moving away from you on Earth, both of which are also c .

The Doppler effect In 1842, Christian Doppler, a professor of mathematics in Prague, deduced that wavelength, hence color, is affected by motion. As objects move toward or away from you, they change color, in analogy to how the pitch of a siren changes as it passes you. As shown for the observer on the left in **Figure 3-5**, the wavelengths of electromagnetic radiation from an approaching source are compressed. The circles represent consecutive wave peaks (S_1 through S_4) emitted in all directions as the source moves along. Because each successive wave is sent out from a position slightly closer to the female observer, she sees a shorter wavelength than she would if the source were stationary. All of the colors in the spectrum of an approaching source, such as a star, are therefore shifted toward the short-wavelength (blue) end of the spectrum, regardless of its distance.

This phenomenon is called a **blueshift**.

Conversely, electromagnetic waves from a receding source are stretched out. The observer on the right in **Figure 3-5** sees a longer wavelength than he would if the source were stationary. All of the colors in the spectrum of a receding source, regardless of its distance,

Focus Question 3-2

Does a police siren approaching you sound higher in pitch (shorter wavelength) or lower in pitch than a siren at rest relative to you?

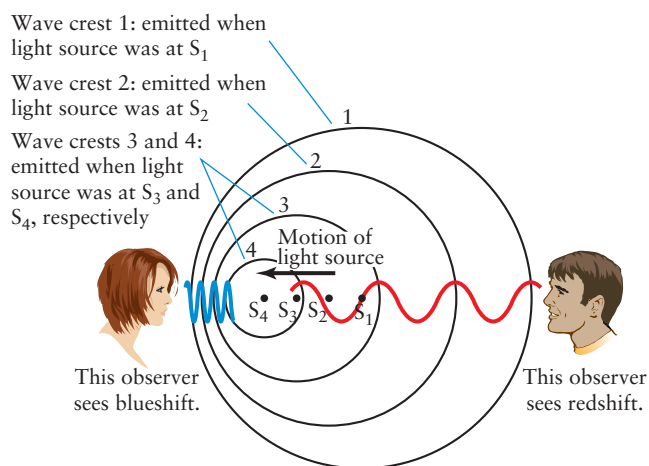


FIGURE 3-5 The Doppler Shift Wavelength is affected by motion between the light source and the observer. A source of light is moving toward the left. The four circles (numbered 1 through 4) indicate the location of the peaks of light waves that were emitted by the moving source when it was at points S_1 through S_4 , respectively. Note that the waves are compressed in front of the source but stretched out behind it. Consequently, wavelengths appear shortened (blueshifted) if the source is moving toward the observer and lengthened (redshifted) if the source is moving away from the observer. Motion perpendicular to the observer's line of sight does not affect wavelength.

are shifted toward the longer-wavelength (red) end of the spectrum, producing a **redshift**. A blueshift or a redshift is also called a **Doppler shift**. The amount of Doppler shift varies directly with approaching or receding speed: When the speeds are small compared to the speed of light, an object that approaches twice as fast as another has all of its colors (or wavelengths) blueshifted twice as much as does the slower-moving object. An object moving away twice as fast as another object has its colors redshifted twice as much as the slower-moving object. Doppler shift applies to spectral lines (discussed shortly) as well.

3-3 Einstein showed that light sometimes behaves as particles that carry energy

1 By 1905, scientists were comfortable with the wave nature of light. However, in that year, Albert Einstein threw a monkey wrench into that theory when he proposed that light is composed of particles that have wave properties, creating what is now called the *wave-particle duality*. He used this idea to explain the *photoelectric effect*. Physicists knew that electrons are bound onto a metal's surface by electric forces and that it takes energy to overcome those forces. Shorter wavelengths of light can knock some electrons off the surfaces of metals, while longer wavelengths of light cannot, no matter how intense the beam of long-wavelength light. Because some colors (or, equivalently, wavelengths) can remove the electrons and others cannot, the electrons must receive different amounts of energy from different colors of light. But how? Einstein proposed that light travels as waves enclosed in discrete packets, now called **photons**, and that photons with different wavelengths have different amounts of energy. Specifically, the shorter the wavelength, the higher a photon's energy.

$$\text{Photon energy} = \frac{\text{Planck's constant} \times \text{the speed of light}}{\text{Wavelength}}$$

where Planck's constant (named for the German physicist Max Planck) has the value $6.67 \times 10^{-34} \text{ J} \cdot \text{s}$, where J is the unit of energy called a *joule*, and the wavelength, the distance between wave crests or troughs, is shown in **Figure 3-3**. ($\text{J} \cdot \text{s}$ stands for joules multiplied by seconds, or *joule-seconds*.) Einstein's concept of light, confirmed in numerous experiments, means that light can act both as waves (as when passing through slits) and as particles (as when striking matter).

The waves shown in **Figure 3-3** are moving to the right. If you count the number of wave crests that pass a given point per second, you have found the **frequency** of the photon. The unit of frequency is the *hertz*, named in honor of the German physicist Heinrich Hertz. One hertz means that one cycle per second—or that one wave crest per second—passes any point. A thousand hertz means a thousand cycles per second, and so on. The frequency is used, among many other things, to

identify radio stations. For example, WCPE radio in Wake Forest, North Carolina, has a frequency of 89.7 megahertz (a megahertz is a million hertz or a million cycles per second).

All photons with the same wavelength are identical to each other and, therefore, every photon of a given wavelength carries the same amount of energy as every other photon with that wavelength. The energy delivered by a photon is either enough to eject an electron from the surface of the metal or it is not; there is no middle ground. Extensive testing in the twentieth century confirmed both the wave and particle properties of light.

While the energy of a single photon is fixed by its wavelength, the total number of photons passing per second from that source with a given energy determines the *intensity* of the electromagnetic radiation at that wavelength (that is, how bright the object appears to be). The more photons detected, the higher the intensity, and vice versa. However, the intensity of light does not change the energy per photon. If one photon is unable to eject an electron from a metal, then billions of photons with that energy will still be unable to remove that electron.

3-4 Visible light is only one type of electromagnetic radiation

We now know that visible light occupies only a tiny fraction of the full range of possible wavelengths, collectively called the **electromagnetic spectrum**. Visible light has the range of wavelengths from 400 to 700 nm. However, Maxwell's equations place no length restrictions on the wavelengths of electromagnetic radiation. What lies on either side of this interval? Around 1800, the British astronomer William Herschel discovered **infrared radiation** in an experiment with a prism. When he held a thermometer just beyond the red end of the visible spectrum, the thermometer registered a temperature increase, indicating that it was being heated by an invisible form of energy (Figure 3-6). Infrared radiation, discovered before Maxwell's equations were formulated, was later identified as electromagnetic radiation with wavelengths slightly longer than red light. Our bodies detect infrared radiation as heat.

In experiments with electric sparks in 1888, Hertz succeeded in producing electromagnetic radiation a few centimeters in wavelength, now known as **radio waves**.

At wavelengths shorter than those of visible light, **ultraviolet (UV) radiation** extends from about 400 nm to 10 nm. In 1895, Wilhelm Roentgen invented a machine that produces electromagnetic radiation with wavelengths shorter than 10 nm, now called **X-rays**. Modern versions of Roentgen's machine are found in medical and dental offices. X-rays have wavelengths between about 10 and 0.01 nm. Photons with even shorter wavelengths are called **gamma rays**. These boundaries are all arbitrary and primarily used as convenient divisions in the electromagnetic spectrum, which is actually continuous.

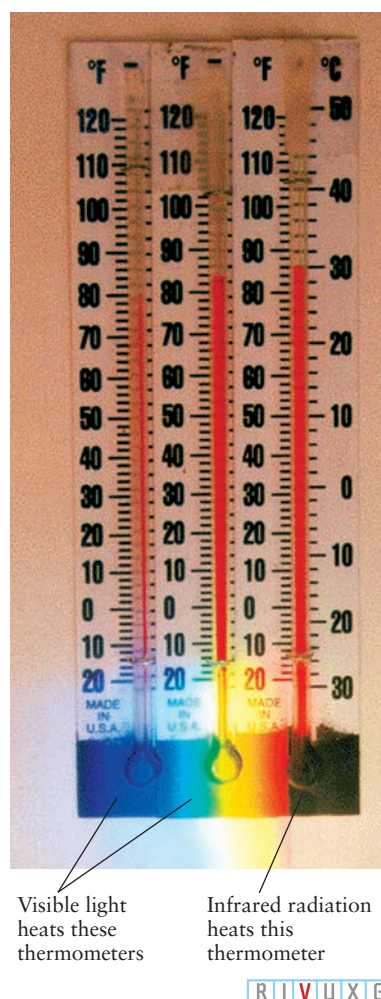


FIGURE 3-6 Experimental Evidence for Infrared Radiation This photograph shows the visible colors separated by a prism. The two thermometers in the region illuminated by visible light have temperatures less than the thermometer to the right of red. Therefore, there must be more radiation energizing (that is, heating) the warmest thermometer. This energy is what we call infrared radiation—invisible to the human eye, but detectable as heat. (NASA/JPL-Caltech)

As shown in Figure 3-7, the electromagnetic spectrum stretches from the longest-wavelength radio waves, through microwaves, infrared radiation, visible light, ultraviolet radiation, and X-rays, to the shortest-wavelength photons, gamma rays. On the long-wavelength side of the visible spectrum, infrared radiation covers the range from about 700 nm to 1 mm. Astronomers interested in infrared radiation often express wavelength in *micrometers* or *microns* (abbreviated μm), where $1 \mu\text{m} = 1000 \text{ nm} = 10^{-6} \text{ m}$ (one millionth of a meter). From roughly 1 mm to 10 cm is the range of microwaves. Microwaves are sometimes considered as a separate class of photons and sometimes categorized as infrared radiation or radio waves. Formally, radio waves are all electromagnetic waves longer than 10 cm.

Focus Question 3-3

Which photon has more energy, an infrared photon or an ultraviolet photon?

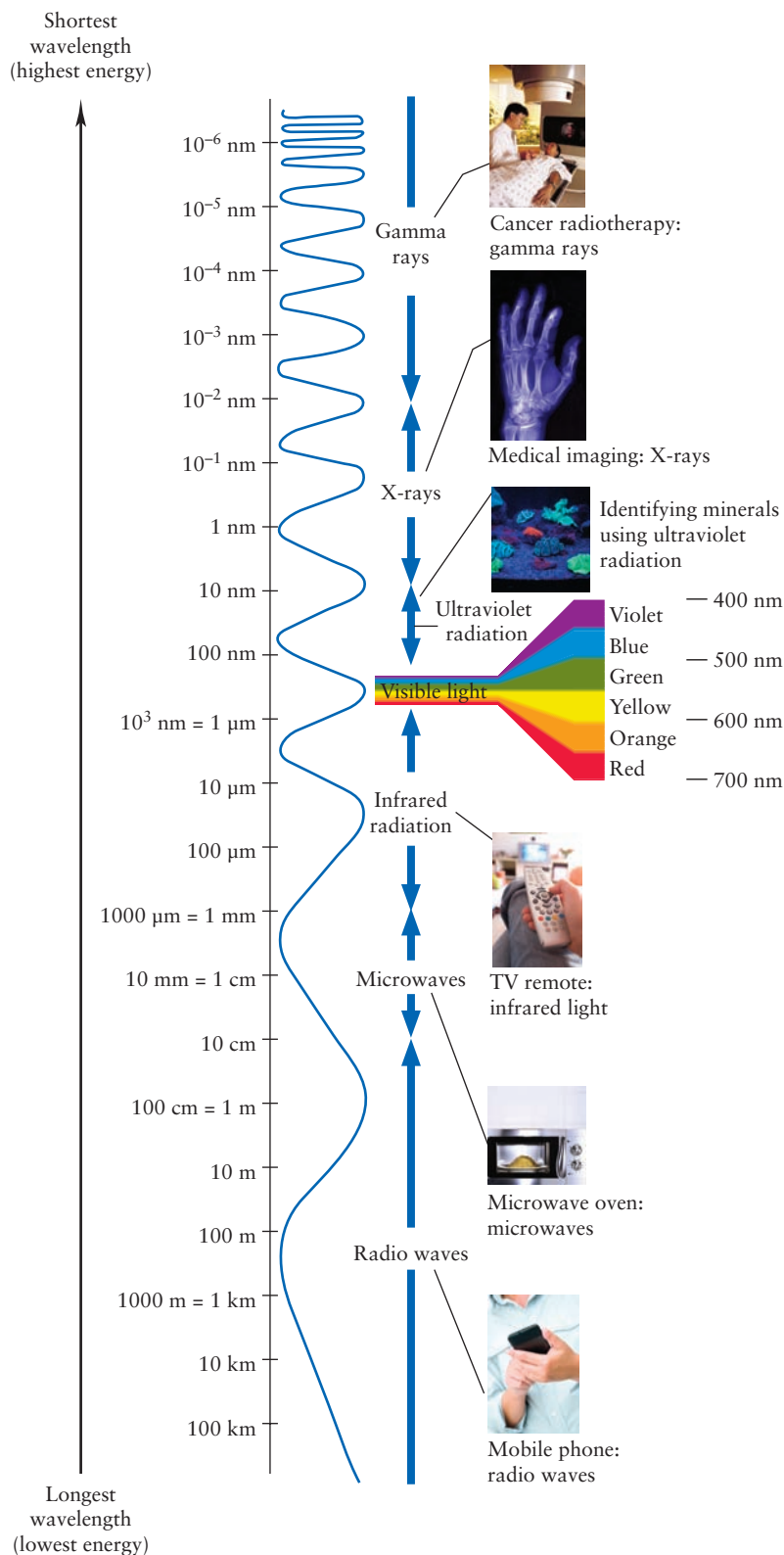


FIGURE 3-7 The Electromagnetic Spectrum The full array of all types of electromagnetic radiation is called the electromagnetic spectrum. It extends from the longest-wavelength radio waves to the shortest-wavelength gamma rays. Visible light forms only a tiny portion of the full electromagnetic spectrum. Note that 1 μm (micrometer) is 10^{-6} m, and 1 nm

(nanometer) is 10^{-9} m. The insets show how we are now using all parts of the electromagnetic spectrum here on Earth. (From top: Will and Deni McIntyre/Science Source; Ted Kinsman/Science Source; Chris Martin Bahr/Science Source; Bart Broek/E+/Getty Images; Michael Haegle/Corbis; iStockphoto/Thinkstock)

The various types of electromagnetic radiation share many basic properties. For example, they are all photons, they all travel at the same speed, and they all sometimes behave as particles and sometimes as waves. But, because of their different wavelengths (and therefore different energies), they interact very differently with matter. For example, X-rays penetrate deeply into your body tissues, while visible light is mostly stopped and scattered by the surface layer of skin; your eyes respond to visible light but not to infrared radiation; and your radio detects radio waves but not ultraviolet radiation.

Earth's atmosphere is relatively transparent to visible light, radio waves, microwaves, short-wavelength infrared, and long-wavelength ultraviolet. As a result, these radiations pass through the atmosphere without much loss and can be detected by ground-based telescopes sensitive to them. Astronomers say that the atmosphere has *windows* for these parts of the electromagnetic spectrum (Figure 3-8).

The longest-wavelength ultraviolet radiation, called UVA, causes tanning and sunburns. Ozone (O_3) in Earth's atmosphere normally screens out intermediate-wavelength ultraviolet radiation, or UVB. Until recently, the ozone in the *ozone layer* high in the atmosphere was being depleted by human-made chemicals, such as chlorofluorocarbons (CFCs) and bromine-rich gases. As a result, more UVB is reaching Earth's surface, and these highly energetic photons severely damage living tissue, causing a surge in the rates of skin cancer and glaucoma, among other diseases.

2 Earth's atmosphere is completely opaque to the other types of electromagnetic radiation, meaning that they do not reach Earth's surface. (This opacity is a good thing, because short-wavelength ultraviolet radiation [UVC], X-rays, and gamma rays are devastating to living tissue. Gamma rays, packing the highest energies per photon, are the deadliest.) Direct observations of these wavelengths must be performed high in the atmosphere or, ideally, from space.

As noted earlier, photons with different energies interact with matter in different ways. Higher-energy photons will pass through or rip apart material from which lower-energy photons will bounce off. Telescope designs for collecting and focusing radiation, therefore, differ depending on the energies or, equivalently, the wavelengths of interest. Knowing the energies of photons and their effects on matter enables astronomers to design telescopes and recording devices sensitive to them. In the next section we will consider the lengths to which astronomers have gone to capture visible and invisible (or nonoptical) electromagnetic radiation.

OPTICS AND TELESCOPES

3 Since the time of Galileo, astronomers have been designing instruments to collect more light than the

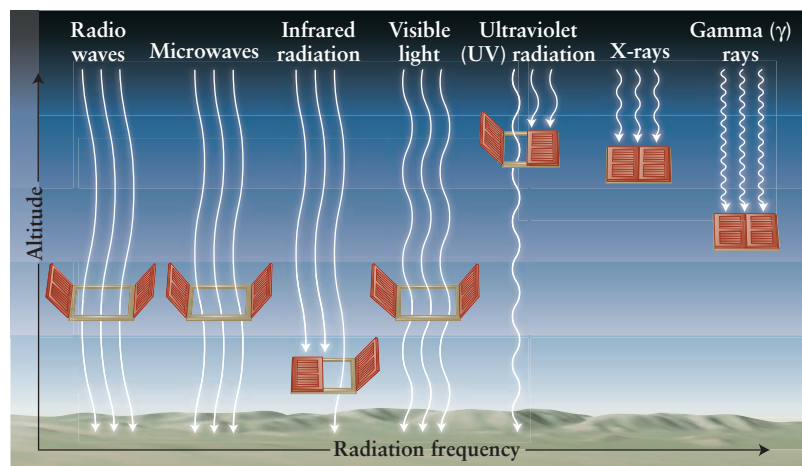


FIGURE 3-8 “Windows” Through the Atmosphere Different types of electromagnetic radiation penetrate into Earth's atmosphere in varying amounts. Visible light, radio waves, microwaves, short-wavelength infrared, and long-wavelength ultraviolet reach all the way to Earth's surface. The other types of radiation are absorbed or scattered by the gases in the air at different characteristic altitudes (indicated by heights of windows). Although the atmosphere does not have actual windows, astronomers use the term to characterize the passage of radiation through it.

human eye can gather on its own. Collecting more light enables us to see things more brightly, to see dimmer objects, and under the right circumstances to magnify objects and to see them in more detail. There are two basic types of telescopes—those that collect light through lenses, or **refracting telescopes**, and those that collect it from mirrors, or **reflecting telescopes (reflectors)**. The earliest telescopes, such as Galileo's, used lenses, which have a variety of shortcomings as light-gathering devices. Consequently, all modern research telescopes use mirrors to collect light. Lenses are still used to gather light in some small home telescopes and as eyepieces that enable us to look directly into any telescope. Lenses are also used to collect light in binoculars and cameras. Because astronomers exclusively use reflecting telescopes, we will begin exploring telescopes by discussing how these telescopes work. Then we will consider how lenses collect light, how refracting telescopes work, and, finally, how astronomers have developed telescopes to see nonvisible electromagnetic radiation.

3-5 Reflecting telescopes use mirrors to concentrate incoming starlight

The first reflecting telescope was built in the seventeenth century by Isaac Newton. To understand how these telescopes work, consider a flow of photons, more commonly called a *light ray*, moving toward a flat mirror. In **Figure 3-9a**, a light ray strikes the mirror, and we imagine a perpendicular line coming out of the mirror at that point. According to the principle of **reflection**, the

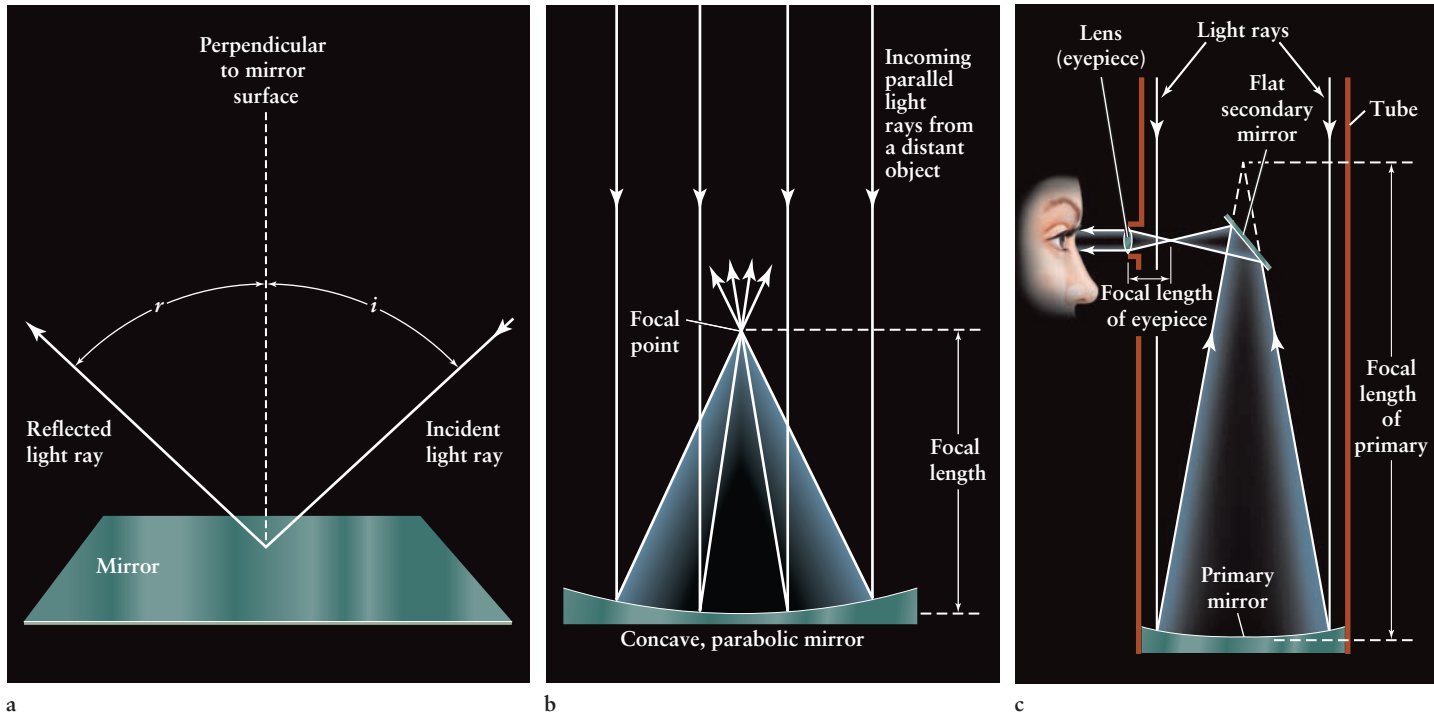


FIGURE 3-9 Reflection (a) The angle at which a beam of light strikes a mirror (the angle of incidence, i) is always equal to the angle at which the beam is reflected from the mirror (the angle of reflection, r). (b) A concave, parabolic mirror causes parallel light rays to converge and meet at the focal point. The distance between the mirror and focal point is the focal length. (c) A Newtonian tele-

scope uses a flat mirror, called the secondary mirror, to send light toward the side of the telescope. The light rays are made parallel again by passing through a lens, called the eyepiece. The dashed line shows where the focal point of this primary mirror would be if the secondary mirror were not in the way.

angle between the incoming light ray and the perpendicular (dashed line) is always equal to the angle between the outgoing, reflected light ray and that perpendicular. This principle is often stated as, “The angle of incidence equals the angle of reflection.” This rule also applies if the mirror is curved (Figure 3-9b).

Using this principle, Newton determined that a concave (hollowed-out) mirror, ground in the shape of a parabola, causes all parallel incoming light rays that strike the mirror to converge to a **focal point** (Figure 3-9b). The distance between this **primary mirror** and the focal point, where the image of the distant object is formed, is called the **focal length** of the mirror. Focal points exist for light from sources that are extremely far away, like the stars. (Figure 3-10 shows why stars can be considered “far away.”) If the object is larger than a point, like the Moon or a planet, then the light will converge to a plane, called the **focal plane**, located at the distance of the focal length. Likewise, point objects like stars that are near each other on the celestial sphere also come into focus near each other on the focal plane.

To view the image, Newton placed a small, flat mirror at a 45° angle between the primary mirror and the focal point, as sketched in Figure 3-9c and Figure 3-11a. This **secondary mirror** reflects the light rays to one side of the telescope, and the viewer observes the image through

an **eyepiece lens**. We will discuss how this lens works in Section 3-7. A telescope with this optical design is still called a **Newtonian reflector**. Research telescopes do not use eyepieces. As we will see shortly, light-sensitive detectors are placed in their focal planes instead.

Newtonian telescopes are popular with amateur astronomers because they are convenient to use while the observer is standing up. However, they are not used in research observatories because they are lopsided. If astronomers attach their often heavy and bulky research equipment onto the side of a Newtonian telescope, the telescope twists and distorts the image in unpredictable ways. Making matters worse, these distortions change as the telescope tracks stars across the sky and therefore tilts at different angles.

Three basic designs exist for the reflecting telescopes used in research: Cassegrain, Nasmyth/coudé, and prime focus telescopes. In the first, a hole is drilled directly through the center of the primary mirror. A convex (outwardly curved) secondary mirror placed between the primary mirror and its focal point reflects the light rays back through the hole (Figure 3-11b). This design is called the **Cassegrain focus**, after a 1672 design by Laurent Cassegrain. This secondary mirror extends the telescope’s focal length. Compact, relatively low-weight equipment is bolted to the bottom of the telescope, and

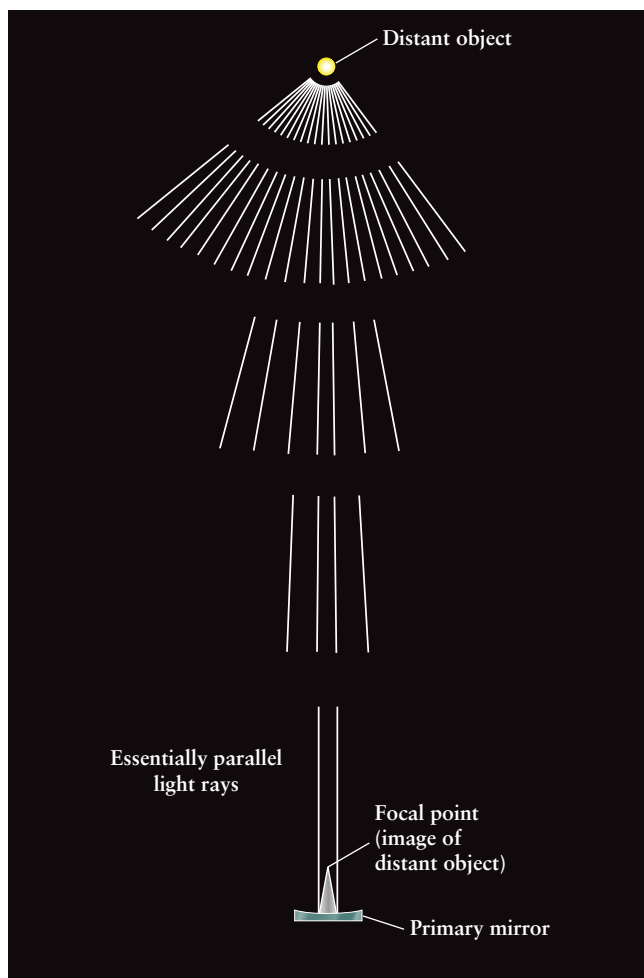


FIGURE 3-10 Parallel Light Rays from Distant Objects As light travels away from any object, the light rays, all moving in straight lines, separate. By the time light has traveled trillions of kilometers, only the light rays moving in virtually parallel tracks are still near each other.

the light is brought into focus in it. The advantage of this design over Newtonian telescopes is that the attached equipment is balanced and does not distort the telescope frame and, hence, the image.

The second design that astronomers use has two variations, both of which use a third mirror to direct light out the side of the telescope, at the place where it pivots (Figure 3-11c). Heavier or bulkier optical equipment that requires firmer mounting can be located at the *Nasmyth focus*, after Scottish engineer James Nasmyth, who developed it. If an extremely long focal length is desired, a *coudé focus* (named after a French word meaning “bent like an elbow”) is used. Equipment that profits from the use of the Nasmyth focus and coudé focus includes a variety of *spectrographs*, instruments that separate light from objects into its individual colors to determine the objects’ chemistries, surface temperatures, and motions toward or away from us.

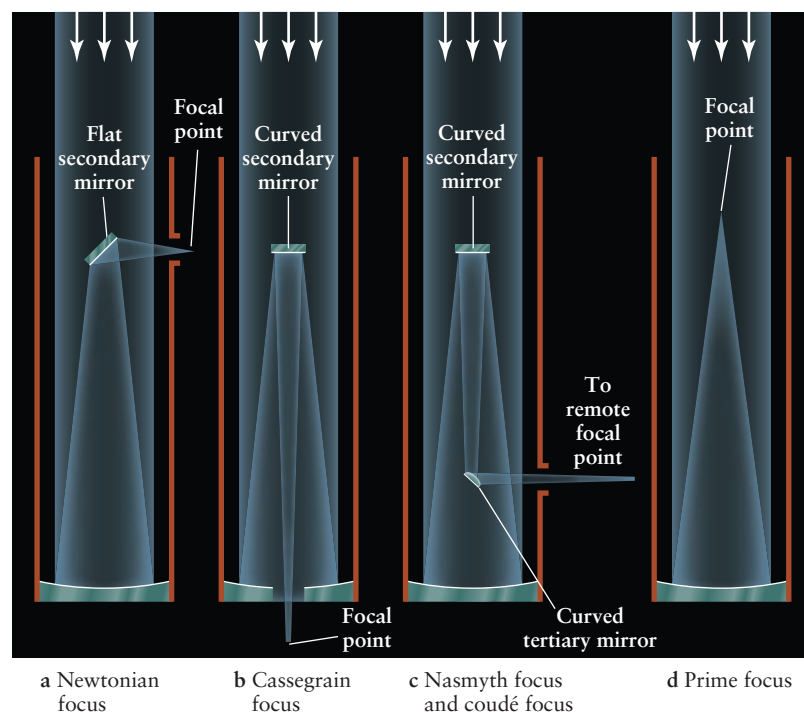


FIGURE 3-11 Reflecting Telescopes Four of the most common optical designs for reflecting telescopes: (a) Newtonian focus (popular among amateur astronomers), and the three major designs used by researchers: (b) Cassegrain focus, (c) Nasmyth focus and coudé focus, and (d) prime focus.

In the third design, an observing device is located at the undeflected focal point, directly in front of the primary mirror. This arrangement is called the **prime focus** (Figure 3-11d). This design has the advantages of making the brightest image (for a given exposure time) and having no secondary mirror, having the fewest reflections that otherwise would cause light loss and distortion. The earliest telescopes were ground from a metal alloy called speculum metal that polished to a bright surface. However, these mirrors reflected only about two-thirds of the light that struck them, they tarnished rapidly, and they changed shape (hence, distorted images) as the air temperature changed. Modern primary telescope mirrors today are made from glass, which maintains its shape well. A highly reflective aluminum coating is applied to the top surface of the mirror after it is ground and polished to the appropriate shape. In this design, the light never enters the mirror, unlike the mirrors we use in everyday life, with the reflective coating behind the glass. Aluminized telescope mirrors typically reflect over 90% of the light that strikes them. The mirrored surfaces of research telescopes are polished so smoothly that the highest bumps are less than 0.002 times the thickness of a human hair.

You have probably noticed (see Figure 3-11) that the secondary mirror of a reflector blocks some incoming light—one unavoidable price that astronomers must

pay. Typically, a secondary mirror prevents about 10% of the incoming light from reaching the primary mirror. This problem is addressed by constructing primary mirrors with sufficiently large surface areas to compensate for the loss of light. You might also think that, because light is missing from the center of the telescope due to blockage by the secondary mirror, a corresponding central “hole” appears in the images. However, this problem does not occur, because light from all parts of each object being observed enters all parts of the telescope (Figure 3-12). Indeed, covering any part of the primary mirror darkens the image but does not limit which parts of the object you can see through the telescope. Likewise, blocking part of the opening of a telescope made just with lenses darkens the image but does not create a hole in the image.

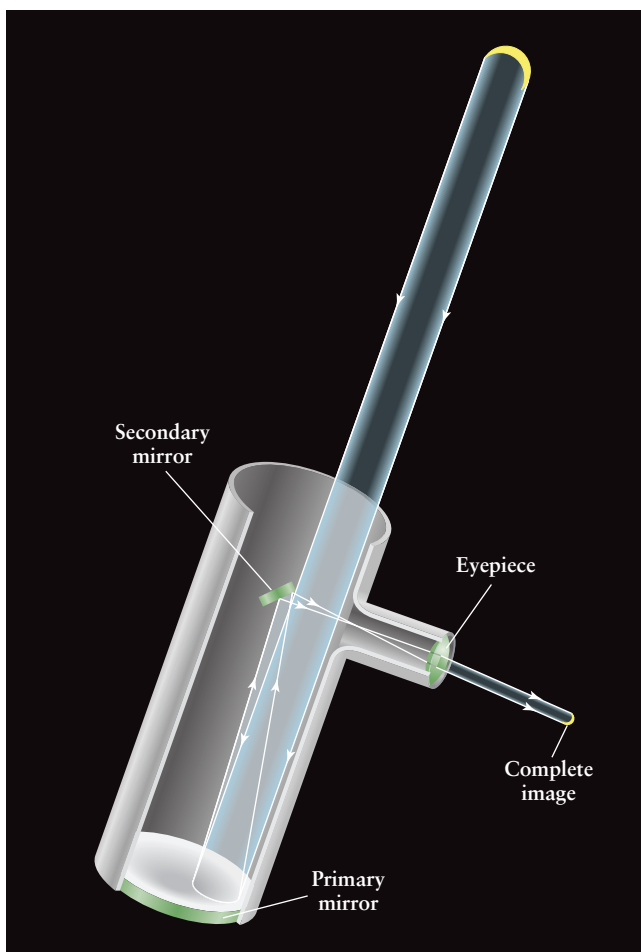


FIGURE 3-12 Secondary Mirror of a Reflecting Telescope Does Not Create a Hole in the Image Because the light rays from distant objects are parallel, light from the entire object (in this case, the crescent Moon) reflects off all parts of the primary mirror. Therefore, every part of the object sends photons to the eyepiece. This figure shows the reconstruction of the entire crescent Moon from light passing through just part of this telescope. The same drawing applies everywhere on the primary mirror that is not blocked by the secondary mirror.

Insight Into Science

Costs and Benefits Science now relies heavily on technology to conduct experiments and to make observations. The cost of cutting-edge astronomical observations may run to hundreds of millions of dollars or more. The return on such investments is a better understanding of how the universe works, how we can harness its resources, and our place in it.

3-6 Telescopes brighten, resolve, and magnify

As mentioned earlier, a telescope’s most important function is to provide astronomers with as bright an image as possible. The brighter an object appears, the more information about it we can extract. The observed brightness of any object depends on the total number of photons collected from it, which, in turn, depends on the area of the telescope’s primary mirror. Analogously, the pupils in our eyes get larger in dark environs to allow more photons to strike the retina and create a brighter image than otherwise. For exposures of equal times, a telescope with a large primary mirror produces brighter images and detects fainter objects than a telescope with a smaller primary mirror (Figure 3-13).

The **light-gathering power** of a telescope is directly related to the area of the telescope’s primary mirror. Recall that the area and diameter of a circle are related by the formula

$$\text{Area} = \frac{\pi d^2}{4}$$

where d is the diameter of the mirror and π (pi) is about 3.14. (You probably know this equation as $\text{Area} = \pi r^2$, where r , the radius, is half the diameter.) Consequently, *a mirror with twice the diameter of another mirror has 4 times the area of the smaller mirror and, therefore, collects 4 times as much light as does the smaller one in the same amount of time.* For example, a 36-cm-diameter mirror has 4 times the area of an 18-cm-diameter mirror. Therefore, the 36-cm telescope has 4 times the light-gathering power of a telescope half its size.

Another vital function of any telescope is to reveal greater detail of objects that are more than just points of light. Such *extended objects* include the Moon, the Sun, the planets in our solar system, galaxies, interstellar gas clouds, and clusters of stars, among other things. A large telescope increases the sharpness of the image and the degree of detail that can be seen. **Angular resolution** (often just called *resolution*) measures the clarity of images (Figure 3-14). The angular resolution of a telescope is measured as the arc angle between two adjacent stars

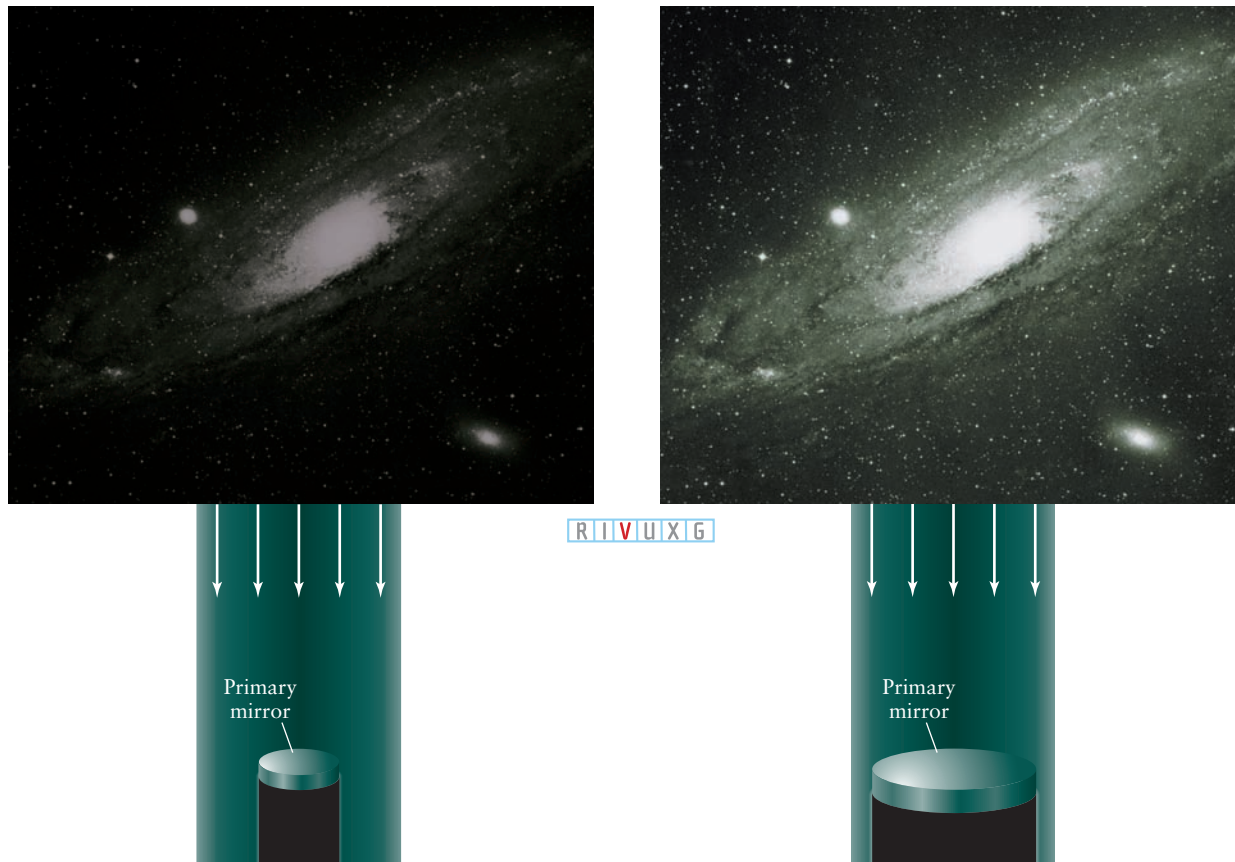
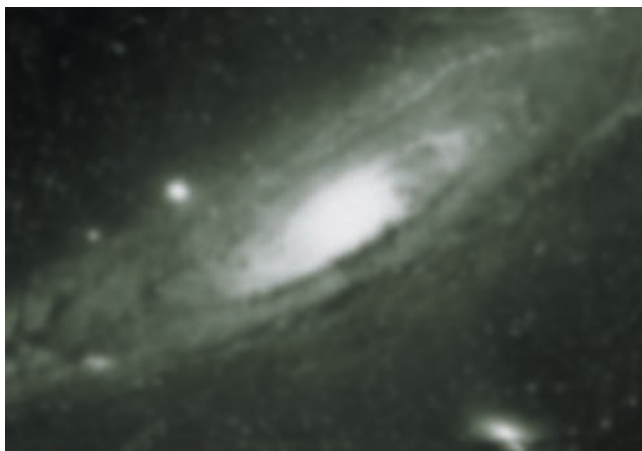


FIGURE 3-13 Light-Gathering Power Because a large primary mirror collects more starlight than does a smaller one, a larger telescope produces a brighter image than a smaller one, all other things being equal. The same principle applies to telescopes

that collect light using just lenses. The two photographs of the Andromeda Galaxy were taken through telescopes with different diameters and were exposed for equal lengths of time at equal magnification. (Royal Observatory, Edinburgh/Science Source)



a

FIGURE 3-14 Resolution The larger the diameter of a telescope's primary mirror, the finer the detail the telescope can resolve. These two images of the Andromeda Galaxy, taken through telescopes with different diameters, show this effect. **(a)** A lower-resolution image taken through a smaller telescope. In this photograph most individual stars blur together to make



b

the galaxies look like fuzzy blobs. **(b)** The same field of view through a larger-diameter telescope. Many more individual stars and interstellar gas clouds are visible here than in (a). Increasing the exposure time of the smaller-diameter telescope (a) will only brighten the image, not improve the resolution. (Royal Observatory, Edinburgh/Science Source)

whose images can just barely be distinguished by the telescope. The smaller the angle is, the sharper the image. Large, modern telescopes, like the Keck telescopes in Hawaii, have angular resolutions better than 0.1 arcsec.

Focus Question 3-4

Normal eyeglasses or contact lenses are designed to improve which aspect of vision presented in this section?

As a general rule, *a telescope with a primary mirror twice the diameter of another telescope's primary provides twice the resolution of the smaller telescope.*

The final function of a telescope is to make objects appear larger. This property is called **magnification**. Magnification is associated with resolution, because the larger the image, the more detail of the image you can potentially see (Figure 3-15). The magnification of a reflecting telescope is equal to the focal length of the primary mirror divided by the focal length of the eyepiece lens:

$$\text{Magnification} = \frac{\text{Focal length of primary}}{\text{Focal length of eyepiece}}$$

For example, if the primary mirror of a telescope has a focal length of 100 cm and the eyepiece has a focal length of 0.5 cm, then the magnifying power of the telescope is

$$\text{Magnification} = \frac{100 \text{ cm}}{0.5 \text{ cm}} = 200$$

This property is usually expressed as 200×

Eyepieces are removable. Eyepieces with different focal lengths change a telescope's magnification. However, there is a limit to the magnification of any telescope. Try to magnify beyond that limit, and the image becomes distorted. As a rule, *a telescope with a primary mirror twice the diameter of another telescope's primary will have twice the maximum magnification of the smaller telescope.*

3-7 Eyepieces, refracting telescopes, and binoculars use lenses to focus incoming light

Although light travels at about 300,000 km/s in a vacuum, it moves more slowly through any medium, such as glass. As light enters the glass, it slows abruptly, much like a person walking from a hard pavement onto a sandy beach. Conversely, light that exits a piece of glass resumes its original speed, just as a person resumes his or her normal pace when stepping back from sand onto a solid surface.

As a result of changing speed, light also changes direction as it passes from one transparent medium into another. As noted earlier, this latter change is called *refraction*. You see refraction every day when looking through windows. Imagine a stream of photons from a star that enters a window, as shown in Figure 3-16a. As a light ray goes from the air into the glass, the light ray's direction changes so that it is more perpendicular to the surface of the glass than it was before entering. Once



a

RITVUXIG



b

RITVUXIG

FIGURE 3-15 Magnification The same telescope can magnify by different amounts, depending on the focal length of the eyepiece. **(a)** A low-magnification image of the Moon. **(b)** An image of the Moon taken with magnification 4 times greater than

image (a). Note in this case that the increased magnification leads to increased resolution (that is, more detail can be seen in the larger image). (iStockphoto/Thinkstock)

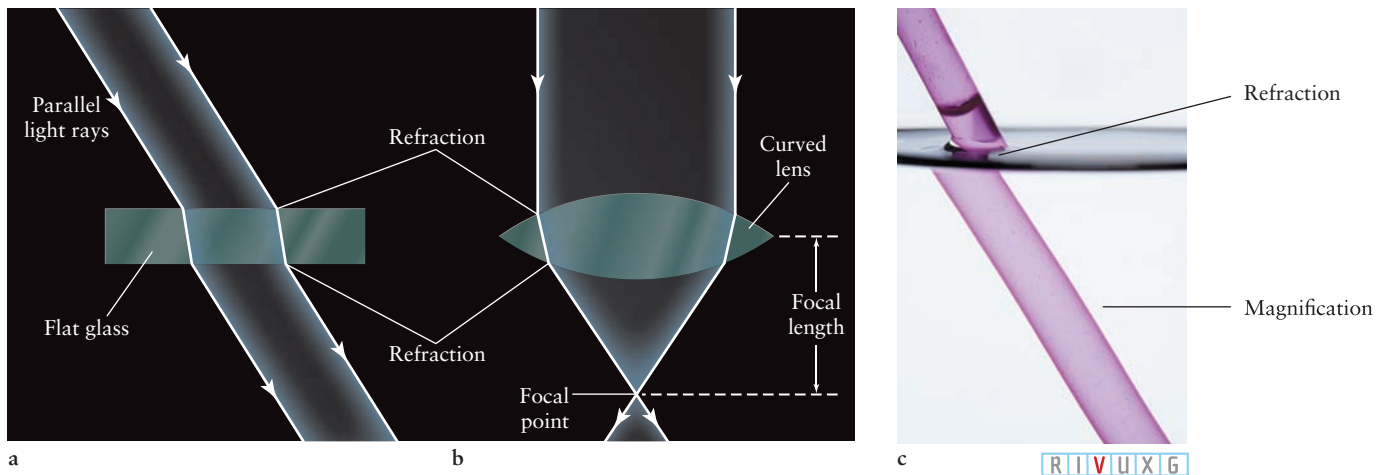


FIGURE 3-16 Refraction Through Uniform and Variable-Thickness Glasses (a) Refraction is the change in direction of a light ray when it passes into or out of a transparent medium such as glass. A light ray that enters a denser medium, such as moving from air into water or glass, is bent or refracted to an angle more perpendicular to the surface than the angle at which it was originally traveling. If the glass is flat, then the light leaving it is refracted back to the direction it had before entering the glass. There is no overall change in the direction in which the light travels. (b) If the glass is in the shape of a

inside the glass, the light ray travels in a straight line. Upon emerging from the other side, the light ray bends once again, resuming its original direction and speed. The net effect is only a slight, uniform displacement of the objects beyond the glass.

Unlike windows, lenses have surfaces of varying thickness (Figure 3-16b). These curved surfaces force the light rays to emerge from the lens in different directions than they had before entering the lens. Lenses that are thicker at their centers than at their edges are called *convex* lenses and force the light that enters them to converge as it passes through. Conversely, lenses thinner at their centers than around their edges are called *concave* lenses. They cause light rays to diverge.

Large convex lenses called **objective lenses** were used instead of primary mirrors in research telescopes in the nineteenth century (and are still used in many telescopes built for home use today). Parallel light rays that enter an objective lens converge and meet at the focal point of the lens. The distance between the focal point and the lens is the lens's focal length (see Figure 3-16b). As with reflecting telescopes, if the object is close enough or large enough to be more than just a dot as seen through the telescope, all the light from it does not converge at the focal point, but rather focuses along the focal plane (Figure 3-17). When we want to look directly into the telescope, a small eyepiece lens is used to bring the light rays collected by the objective lens (or primary mirror) back to being parallel, so our brains can make sense of what we see.

A refracting telescope or **refractor** (Figure 3-18) is therefore an arrangement of two or more lenses used

suitable convex lens, parallel light rays converge to a focus at the focal point. As with parabolic mirrors, the distance from the lens to the focal point is called the focal length of the lens. (c) The straw as seen through the side of the liquid is magnified and offset from the straw above the liquid because the liquid is given a curved shape by the side of the glass. The straw, as seen through the top of the liquid, is refracted but does not appear magnified because the surface of the water is flat and the beaker has uniform thickness. (c: cheyennezj/Shutterstock)

to gather light. The objective lens at the top of the telescope has a large diameter and long focal length. Like a primary mirror, its purpose is to collect as much light as possible. The eyepiece lens, at the bottom of the telescope, is smaller and has a short focal length. The mathematics of magnification for a refracting telescope is the same as for a reflecting telescope with the focal length of the objective lens replacing the focal length of the primary mirror. Likewise, all of the rules for the limits on telescopes are the same for refractors as for reflectors.

The largest refracting telescope in the world, completed in 1897 and located at the Yerkes Observatory in Williams Bay, Wisconsin, near Chicago (Figure 3-19), has an objective lens that is 102 cm (40 in.) in diameter with a focal length of $19\frac{1}{3}$ m ($63\frac{1}{2}$ ft). The lens was ground by the premier American lens-making firm of the nineteenth century, Alvan Clark and Sons. The second-largest refracting telescope, located at Lick Observatory near San Jose, California, has an objective lens of 91 cm (36 in.) in diameter. No major refracting telescopes were constructed in the twentieth century or are planned for this century.

Refracting telescopes suffer from a variety of problems that have limited their use as research instruments. These problems include the following:

- It is difficult to grind a lens to the very complicated shape necessary to have all parallel light rays of any color passing through it converge to the same focal point. When, for example, the lens is given a spherical shape (relatively easy to manufacture), the light at different

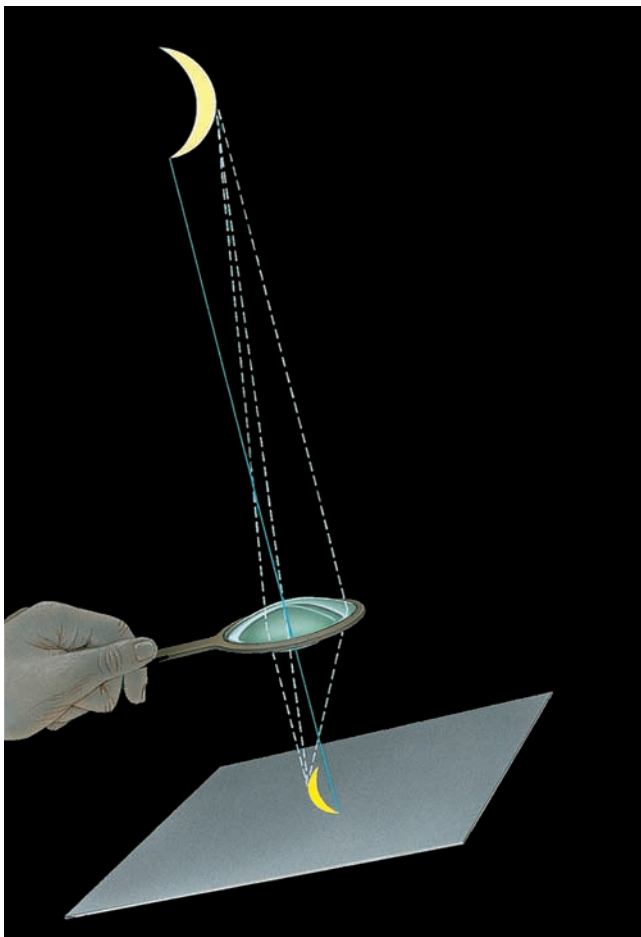


FIGURE 3-17 Extended Objects Create a Focal Plane Light from objects larger than points in the sky does not all converge to the focal point of a lens. Rather, an image of the object is created at the focal length in what is called the focal plane.

distances from the center of the lens has different focal lengths (Figure 3-20). This phenomenon is called **spherical aberration**. Such aberration does not occur for reflecting telescopes because the shape of a mirror needed to focus light from every part of the mirror to the same focal point is a simple parabola, which is easy to manufacture, as discussed in Section 3-8.

Focus Question 3-5

How do human eyes focus light?

- Even when light passes through an ideally shaped lens, different colors of light are refracted by different amounts, so they have different focal lengths. This phenomenon is called **chromatic aberration** (Figure 3-21a). As a result of spherical and chromatic aberration, objects look blurred (Figure 3-21b). Chromatic and spherical aberration can be corrected (Figure 3-22) by adding a second objective lens with a different shape and made of a different type of glass (which refracts light by a different amount than does the first objective lens). Chromatic aberration is not a problem for reflectors because the reflecting surface is always on the top of the mirror; reflecting telescopes do not use refraction to focus light.
- A lens must be supported only around its edges to avoid blocking the light. The weight of a large lens can cause it to sag and thus distort the image. This distortion is not a problem with reflectors because the entire underside of the mirror can be supported, as necessary.
- Air bubbles in the glass cause unwanted refractions and, hence, distorted images.
- Glass does not allow all wavelengths to pass through it equally.

4 These last two points are not issues with reflectors because the light never enters the glass. The combination of all the problems listed here makes images from refractors less accurate than those obtained from reflecting telescopes, which explains why modern research telescopes are all reflectors.

Although they are better light collectors than refracting telescopes, reflecting telescopes are not perfect; there are several prices to pay for the advantages they offer over refractors. Two of the most important prices to pay are blocked light, as discussed in Section 3-5, and spherical aberration, which we consider next.

3-8 Shaping telescope mirrors and lenses is an evolving science

To make a reflector, an optician traditionally grinds and polishes a large slab of glass into a concave, spherical surface. Before computer control, grinding a spherical surface was much easier than grinding the ideal parabolic surface. However, light that reflects off a

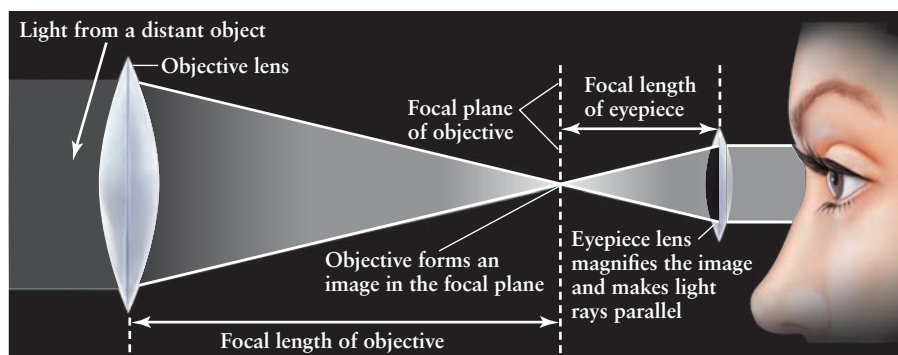


FIGURE 3-18 Essentials of a Refracting Telescope A refracting telescope consists of a large, long-focal-length objective lens that collects and focuses light rays and a small, short-focal-length eyepiece lens that restraightens the light rays. The lenses work together to brighten, resolve, and magnify the image formed at the focal plane of the objective lens.

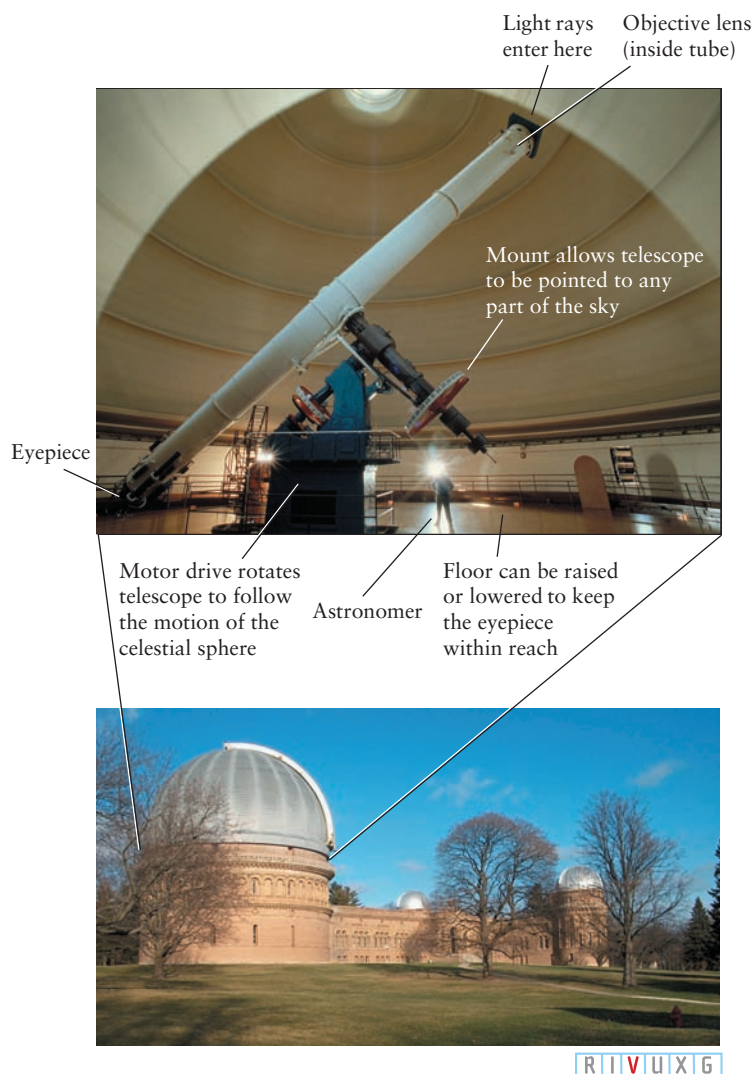


FIGURE 3-19 The Largest Refracting Telescope This giant refracting telescope, built in the late 1800s, is housed at Yerkes Observatory near Chicago. The objective lens is 102 cm (40 in.) in diameter, and the telescope tube is $19\frac{1}{3}$ m ($63\frac{1}{2}$ ft) long. (top: Roger Ressmeyer/Corbis; bottom: Kyle Cudworth)

spherical telescope mirror at different distances from the mirror's center comes into focus at different focal lengths (Figure 3-23a). Images taken directly from such telescopes have the same spherical aberration as the refractors described above.

We can avoid spherical aberration in reflecting telescopes by making the mirror parabolic (Figure 3-23b) or by using a thin correcting lens, called a **Schmidt corrector plate**, with a spherical mirror. Developed in 1930 by the Estonian-Swedish optician Bernhard Schmidt, the corrector plate is located at the top of the telescope (Figure 3-23c). The light coming into the telescope is refracted by the plate just enough to compensate for spherical aberration and to bring all of the light into focus at the same focal length. These correctors have the added benefit of focusing light from a larger angle in the sky than would be in focus without the plate. A Schmidt corrector plate

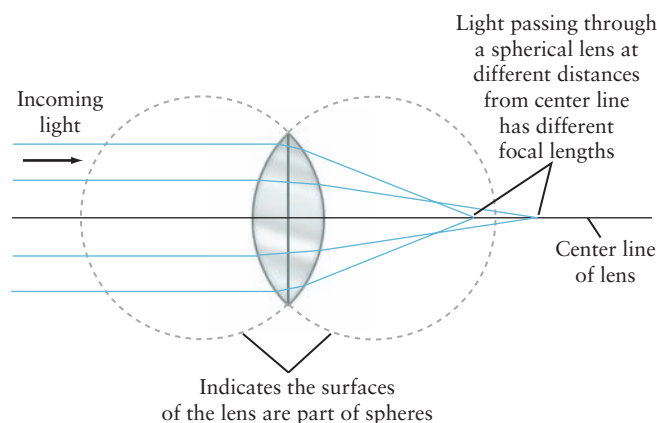
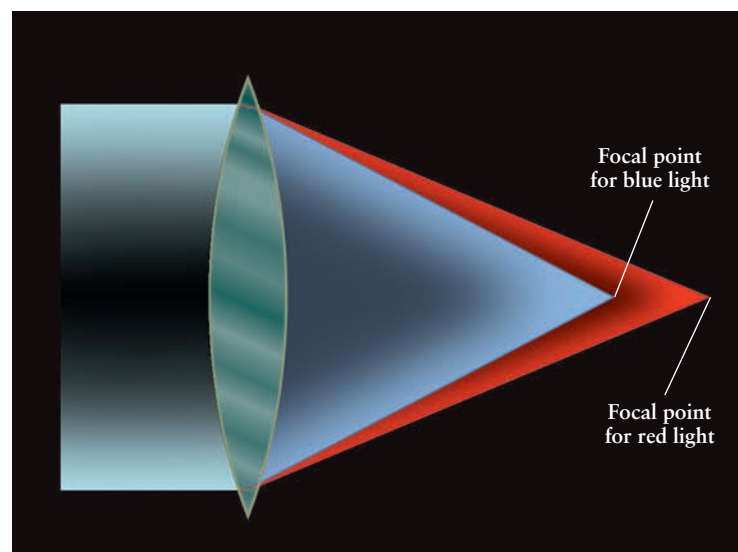


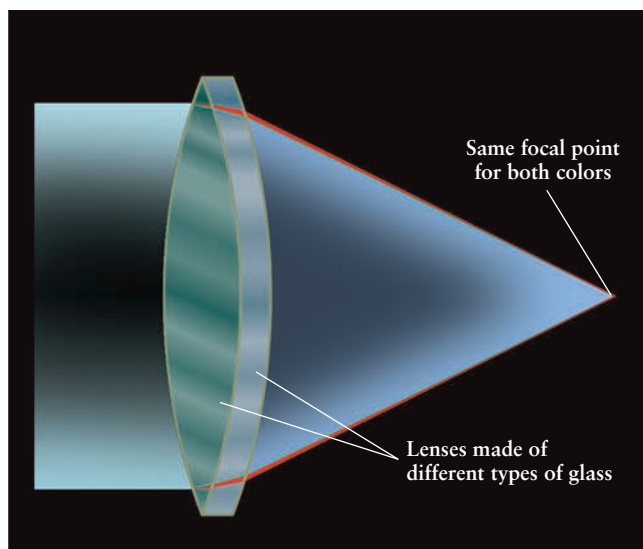
FIGURE 3-20 The Geometry of a Spherical Lens If both sides of a lens are spherical surfaces, as shown here, then light rays of the same color passing through at different distances from the center of the lens are refracted by different amounts. Therefore, spherical lenses have different focal lengths for these different light rays and so they give blurry images.



a The problem: Chromatic aberration



FIGURE 3-21 Chromatic Aberration (a) Light of different wavelengths is refracted by different amounts when passing through a medium such as glass. Therefore, single lenses such as this one have different focal lengths for light of different colors passing through them. (b) Image showing chromatic aberration. Note the different colors on the edges of the petals caused by light passing through a lens. (b: GYRO PHOTOGRAPHY/amanaimagesRF/Getty Images)



a The solution: Use two lenses

FIGURE 3-22 Achromatic Lens (a) By using two differently shaped lenses (often of different types of glass), light of different wavelengths can be brought into focus at the same focal length. Such achromatic lenses are used in cameras and many tele-



b

scopes. (b) Same object as in Figure 3-21b imaged through an achromatic lens. Note that the colors on the edges of the petals seen in Figure 3-21b do not occur here. (b: GYRO PHOTOGRAPHY/amanaimagesRF/Getty Images)

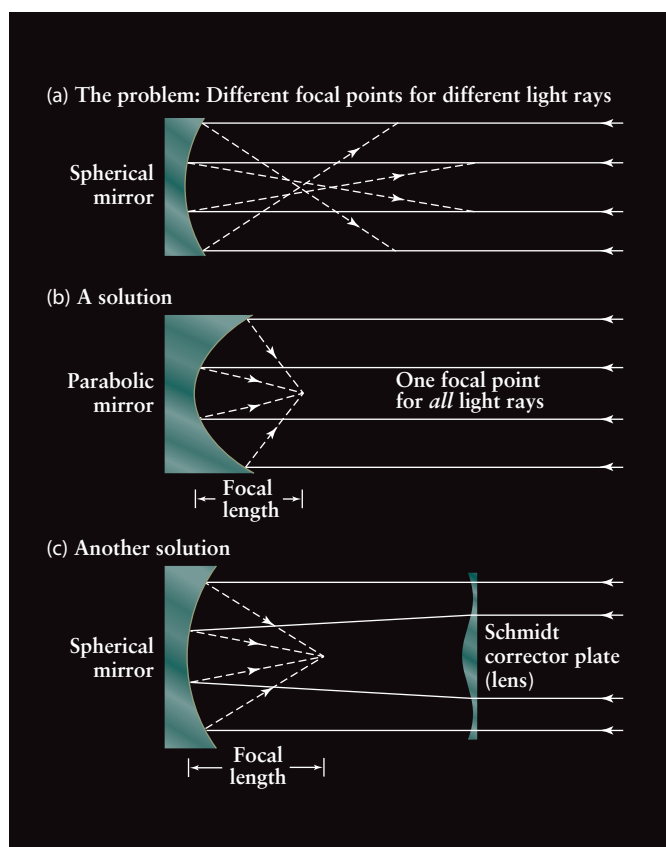


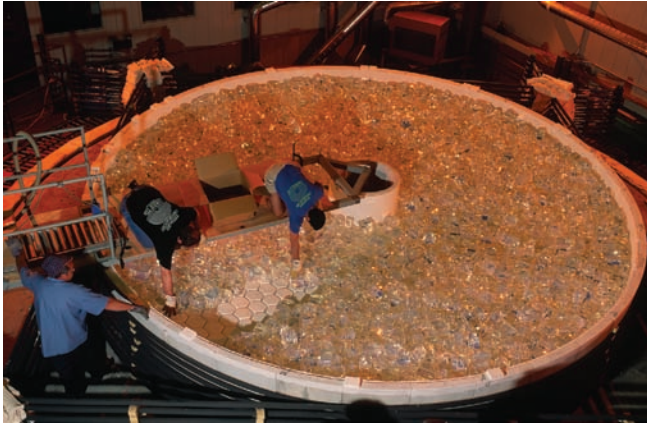
FIGURE 3-23 Spherical Aberration (a) Different parts of a spherically concave mirror reflect light to slightly different focal points. This effect, spherical aberration, causes image blurring. This problem can be overcome by (b) using a parabolic mirror or (c) using a Schmidt corrector plate (a specially curved lens) in front of the telescope.

enables astronomers to map large areas of the sky with relatively few photographs at moderately high magnification. In other words, the Schmidt corrector plate acts like a wide-angle lens on a camera. Schmidt corrector plates are often used in conjunction with Cassegrain mirrors to create *Schmidt-Cassegrain* telescopes, which are popular for home use because they are compact and give relatively wide-angled images. However, the plate does not allow for as much magnification as a telescope with a parabolic mirror.

Although some parabolic primary mirrors have been meticulously hand-ground over the past century, the advent of computer-controlled grinding and then of rotating furnaces (Figure 3-24), in which the liquid glass is actually spun into a parabolic shape, have now made it economical to cast parabolic mirrors with diameters of several meters. That spinning a liquid causes it to develop a parabolic surface was discovered by Isaac Newton in 1689, when he created the effect in a bucket of spinning water.

3-9 Storing and analyzing light from space is key to understanding the cosmos

The invention of photography during the nineteenth century stimulated the development of modern astronomy. The human eye clears itself about 10 times each second. In other words, we store images in our brains for only about 0.1 second. Film, on the other hand, accumulates light as long as it is exposed. By taking a long exposure with a camera mounted at the focal plane of a telescope, an astronomer can record extremely faint features on film that cannot be seen by just looking through the telescope. By keeping telescopes aimed precisely, so that



a

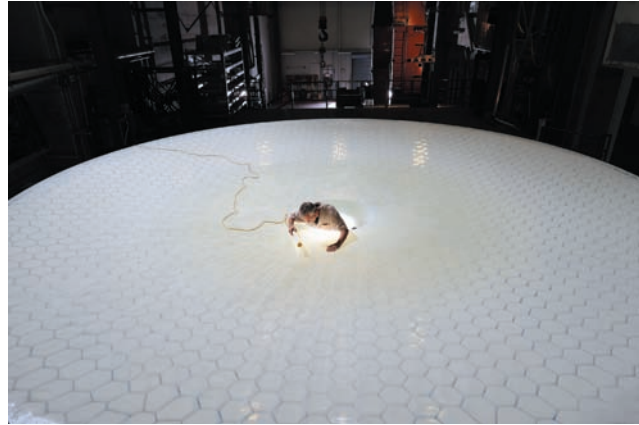
R I V U X G

FIGURE 3-24 Rotating Furnace for Making Parabolic Telescope Mirrors (a) To make each 8.4-m primary mirror for the Large Binocular Telescope II on Mount Graham in Arizona, 40,000 pounds of glass are loaded into a rotating furnace and heated to 1450 K (2150°F). This image shows glass fragments

images do not blur, exposures of an hour or more are now quite routine.

Astronomers have long known, however, that a photographic plate is an inefficient detector of light because it depends on a chemical reaction to produce an image. Typically, only 2% of the light striking film triggers a reaction in the photosensitive material. Thus, roughly 98% of the light falling onto a photographic plate is wasted.

Rapidly evolving technology has changed all that. We have replaced photographic film with highly efficient electronic light detectors called **charge-coupled devices** (CCDs). Very good CCDs respond to between



b

R I V U X G

loaded into the cylindrical furnace. (b) After melting, spinning, and cooling, the mirror's parabolic surface is ready for final smoothing and coating with a highly reflective material. (a: Roger Ressmeyer/Corbis; b: Joe McNally/Getty Images)

50% and 75% of the light falling on them; great CCDs are sensitive to more than 90% of the photons that strike them. Clearly CCD resolution is much better than that of film, and CCDs respond more uniformly to light of different colors. Divided into an array of small, light-sensitive squares called picture elements or, more commonly, **pixels**, each CCD is a square or rectangle a few centimeters on a side. Several of them are often used together to create a higher-resolution image (Figure 3-25). Digital cameras and picture-taking cell phones use this same CCD technology.

Focus Question 3-6

Why do the human eye and brain clear the images that they receive many times per second?



a

R I V U X G

FIGURE 3-25 Mosaic of Charge-Coupled Devices (CCDs) (a) These 40 CCDs combine to provide 378 million light-sensitive pixels that store images collected by the Canada-France-Hawaii Telescope on the dormant volcano Mauna Kea in Hawaii. Electronic circuits transfer the data to a waiting computer. (b) This image of the Rosette Nebula, a region of star formation



b

R I V U X G

5000 ly away in the constellation Monoceros (the Unicorn), was taken with the CCD in (a). The image shows the incredible detail that can be recorded by large telescopes and high-resolution CCDs. (a: J. C. Cuillandre/Canada-France-Hawaii Telescope; b: CFHT/Science Source)

The largest grouping of CCDs used on a telescope has 400 megapixels (a megapixel is 1 million pixels), compared to most digital cameras, which typically have between 12 and 24 megapixels.

When an image from a telescope is focused on the CCD, an electric charge builds up in each pixel in direct proportion to the intensity of the light (number of photons) falling on that pixel. When the exposure is finished, the charge on each pixel is read into a computer. **Figure 3-26** shows one photograph and two CCD images of the same region of the sky, all taken with the same telescope. You can see how many details visible in the CCD images are absent in the ordinary photograph.

3-10 Earth's atmosphere hinders astronomical research

Earth's atmosphere affects the light from objects in space before it reaches ground-based telescopes. Even for those wavelengths that have windows through the atmosphere (see Section 3-4), some light is absorbed in the atmosphere or reflected by it back into space, making all astronomical objects appear dimmer than they appear from space. Also, the air is turbulent and filled with varying amounts of impurities and moisture. You have probably seen turbulence while driving in a car on a hot day, when the road ahead appears to shimmer. Blobs of air, heated by Earth, move upward to create this effect. Light passing through such a blob is refracted, because each hot air mass has a different density than the cooler air around it. Because each blob behaves like a lens, images of objects beyond them appear distorted.

5 The atmosphere over our heads is similarly moving and changing density, and the starlight passing through it is similarly refracted. Because air density changes rapidly, the resulting changes in refraction make the stars appear to change brightness and position rapidly, an effect we see as **twinkling**. When photographed from Earth for more than a few seconds through a traditional reflecting or refracting telescope as described previously in this chapter, twinkling smears out a star's image, causing it to look like a fuzzy blob rather than a pinpoint of light (**Figure 3-27a**). Astronomers use the expression “seeing” to describe how steady the atmosphere is; when the seeing is bad, much twinkling is occurring and, therefore, telescopic images are spread out and blurry.

The angular diameter of a star's smeared-out image, called the **seeing disk**, is a realistic measure of the best possible resolution. The size of the seeing disk varies from one observatory site to another. At Palomar Observatory in California, the seeing disk is roughly 1 arcsec (1"). The best conditions on Earth have been recorded at the observatory on the 14,000-ft summit of Mauna Kea, the tallest volcano on the island of Hawaii. The seeing disk there has been as small as 0.2".

Without the effects of Earth's atmosphere, stars do not twinkle. As a result, photographs taken from telescopes in space reveal stars as much finer points (**Figure 3-27b**), and extended objects, such as planets and galaxies, appear in greater detail.

Light pollution, light that comes from nonastronomical sources, poses another problem for Earth-based telescopes (**Figure 3-28**). Keep in mind that the larger the primary mirror, the more light it gathers and, there-

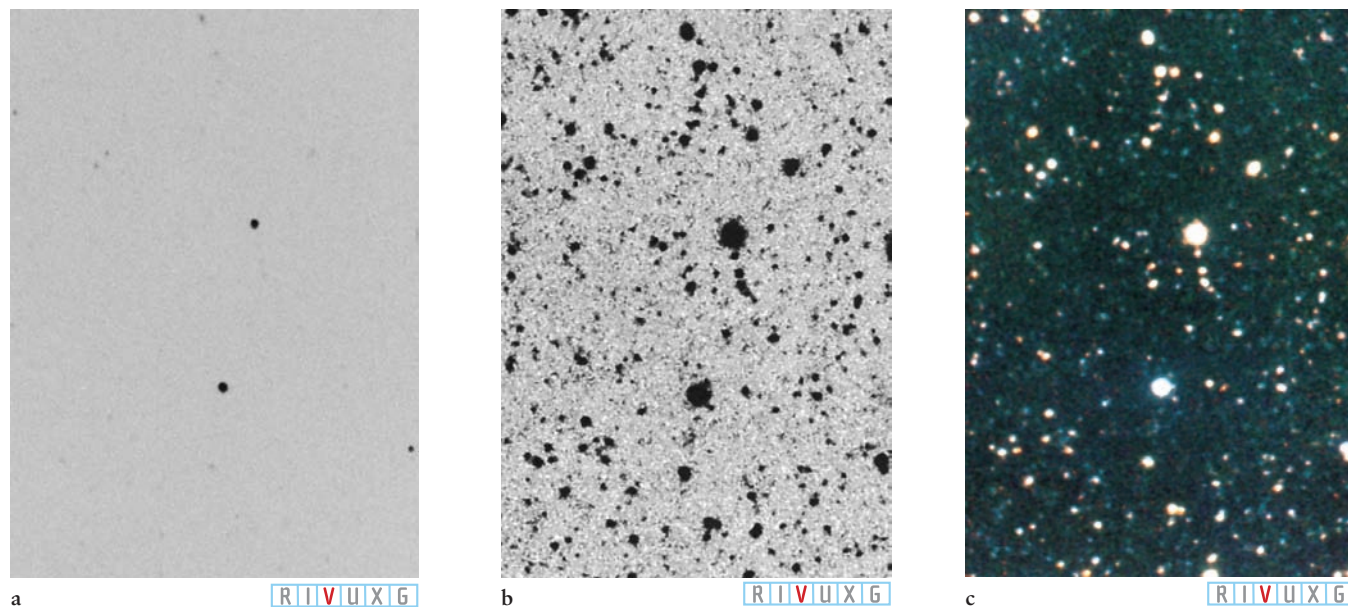
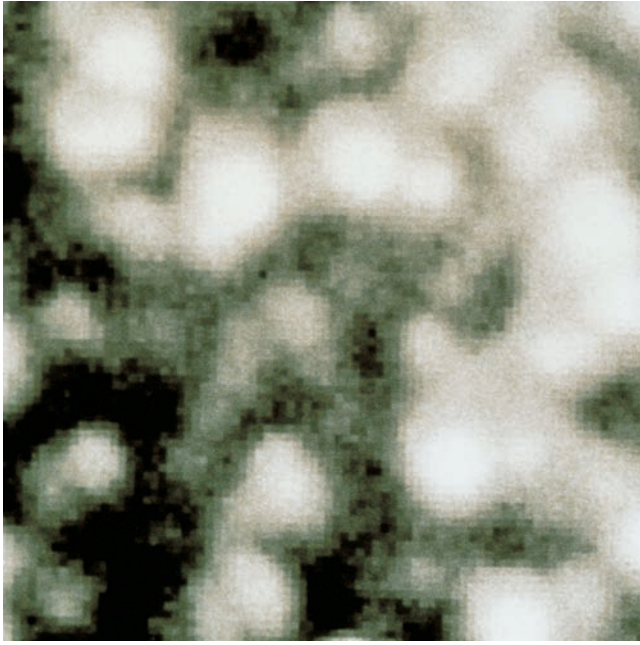


FIGURE 3-26 Photography versus CCD Images These three views of the same part of the sky, each taken with the same 4-m telescope and for the same exposure times, compare CCDs to photographic plates. **(a)** A negative print (black stars and white sky) of a photographic image. **(b)** A negative CCD

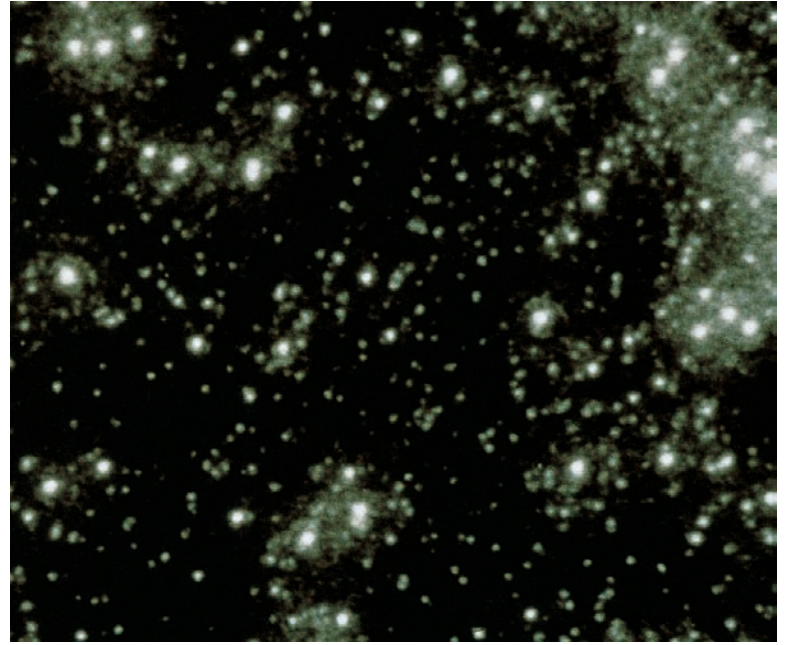
image. Notice that many faint stars that are invisible in the ordinary photograph can be seen clearly in this CCD image. **(c)** This (positive) color view was produced by combining a series of CCD images taken through colored filters. (Patrick Seitzer, NOAO)



a

RIVUXG

FIGURE 3-27 Effects of Twinkling The same star field photographed with (a) a traditional ground-based telescope, which is subject to poor seeing conditions that result in stars twin-



b

RIVUXG

klung, and (b) the Hubble Space Telescope, which is free from the effects of twinkling. (NASA/ESA)

fore, the more information astronomers can obtain from its images or spectra. The 5-m (200-in.) telescope at the Palomar Observatory between San Diego and Los Angeles was the first truly great large telescope, providing astronomers with invaluable insights into the universe for decades. However, light pollution from the two cities now fills the night sky, seriously reducing the ability of

that telescope to collect light from objects in space. Not surprisingly, the best observing sites in the world are high on mountaintops—above smog, water vapor, and clouds—and far from city lights. An even better location, astronomers have discovered, is to observe space from space,

Focus Question 3-7

Where in a typical house can the seeing problems described in this section be observed?



a

RIVUXG

FIGURE 3-28 Light Pollution These two images of New York City, as seen from under the Brooklyn Bridge, show the increase of light in the sky from 1955 to 2010. Compare the skies just above the buildings. This light prevents New Yorkers from seeing dimmer stars that are visible in darker locations.



b

RIVUXG

Since 1972, light pollution, a problem for many observatories around the world, has been partially controlled by local ordinances passed by cities. (top: © D. Corson/ClassicStock/Corbis; bottom: Anghemus Photography/Flickr/Getty Images)

Insight Into Science

Research Requires Patience Seeing conditions—indeed, most observing situations in science—are rarely ideal. Besides such natural phenomena, which are beyond their control, scientists must also contend with equipment failures, late deliveries of parts, and design flaws. Furthermore, because travel time is so long, some missions (like the robotic spacecraft roving on Mars or the *New Horizons* voyage to Pluto and the Kuiper belt) take years or even decades to complete.

eliminating the interference of both the lights of civilization and Earth's atmosphere. Removing these effects is why the Hubble Space Telescope and other orbiting telescopes achieve such magnificent resolution in the images they take.

The Hubble Space Telescope provides stunning details about the universe For decades, astronomers dreamed of observatories in space. Such facilities would eliminate the image distortion created by twinkling and by poor atmospheric transparency due to pollution, volcanic debris, and water vapor. These telescopes could operate 24/7 and over a wide range of wavelengths—from the infrared through the visible range and far out into the gamma-ray part of the spectrum. Over the past four decades, NASA and other space agencies have launched a variety of space telescopes, including four of what NASA calls its Great Observatories. The first Great Observatory to go up was the Hubble Space Telescope (HST).

Soon after HST was placed in orbit from the space shuttle in 1990, astronomers discovered that the telescope's 2.4-m primary mirror had been ground to the incorrect shape. Therefore, it suffered from spherical aberration, which caused its images to be surrounded by a hazy glow. During a repair mission in December 1993, astronauts installed corrective optics that eliminated the problem (Figure 3-29). The telescope was further upgraded in 1997, 1999, 2002, and 2009. Now HST has a resolution of better than 0.1", which is better than can be obtained by telescopes on Earth's surface without the use of advanced technology, as described next.

The observations taken by HST continue to stagger the imagination. Hubble has made discoveries related to the planets in our solar system, planetary systems around other stars, other galaxies and the distances to them, black holes, quasars, the formation of the earliest galaxies, the age of the universe, and many other topics.

Advanced technology is spawning a new generation of superb ground-based telescopes The clarity of images taken by HST and the "seeing" issues described earlier may suggest that ground-based observational astronomy

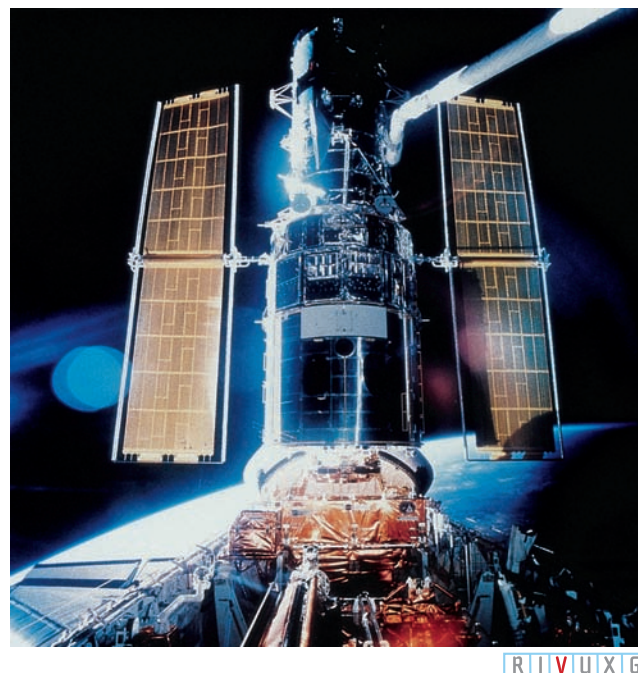


FIGURE 3-29 The Hubble Space Telescope (HST) This photograph of HST hovering above the space shuttle's cargo bay was taken in 1993, at completion of the first servicing mission. HST studies the heavens at infrared, visible light, and ultraviolet wavelengths. (NASA)

is a dying practice. However, two exciting techniques, called *active optics* and *adaptive optics*, enable telescopes on the ground to match the quality of HST—or better it!

Historically, primary mirrors have been thick and heavy, to help keep them rigid and perfectly shaped. Nevertheless, these mirrors were not ideal because they would warp slightly as the telescope changed angle on the sky and the mirrors could not compensate for the seeing conditions, which causes the objects to move and therefore blur. However, in the 1980s, astronomers discovered that changing the shape of the primary mirror would enable them to compensate for these effects. By making thinner mirrors, the shape of the primary could be changed by pistons called *actuators* located under the mirror. Thus, the field of **active optics** was created. This technology finds the best orientation for the primary mirror in response to changes in temperature and the shape of the telescope mount. Actuators adjust the mirror every few seconds to help keep the telescope optimally aimed at its target. With active optics, the New Technology Telescope in Chile and the Keck telescopes in Hawaii routinely achieve resolutions as fine as 0.3", whereas the resolution is much worse for telescopes without active optics at the same sites.

Even better resolution can be achieved with **adaptive optics**, which uses sensors to determine the amount of

twinkling created by atmospheric turbulence. The stellar motion due to this twinkling is neutralized by computer-activated, motorized supports that actually reshape a smaller mirror installed farther down the optical path of the telescope. Adaptive optics effectively eliminates atmospheric distortion and produces remarkably clear images with resolution as fine as 0.03" (Figure 3-30). Many large ground-based telescopes now use adaptive optics on many observing runs, resulting in images comparable to those from HST (Figure 3-30c).

Until the 1980s, telescopes with primary mirrors of between 2 and 6 m were the largest in the world. Now, new technologies in mirror building and computer control allow us to construct much larger telescopes. At least 85 reflectors around the world today have primary mirrors measuring 2 m or more in diameter. Among these are 14 reflectors that have mirrors between 8 and 10.4 m in diameter, with at least 8 other very large telescopes under construction. Appendix S: Largest Optical Telescopes in the World contains a list of the telescopes 3 m in diameter and larger that are in operation.

Because the cost of building very large mirrors is enormous, astronomers have devised less expensive ways to collect the same amount of light. One approach is to make a large mirror out of smaller pieces, fitted together like floor tiles. The largest examples of this segmented-mirror technique are the 10.4-m Gran Telescopio Canarias in the Canary Islands and the 10-m (400-in.) Keck I and Keck II telescopes on the summit of Mauna Kea in Hawaii. In each of these telescopes, 36 curved hexagonal mirrors are mounted side by side to collect the same amount of light as a single primary mirror of 10.4 or 10 m, respectively (Figure 3-31). An-

other method, called **interferometry**, combines images from different telescopes. For example, used together to observe the same object, the two Keck telescopes have the resolving power of a single 85-m telescope. The four 8.2-m reflectors of the Very Large Telescope at the Paranal Observatory (see the photo at the beginning of this chapter) combine to create the resolution that would come from a single telescope 200 m in diameter. This observatory has an ideal resolution of about 0.002".

Focus Question 3-8

Why are military and spy agencies around the world so interested in cutting-edge telescope technologies?

Hint: It is not to look for alien spacecraft.

CAPTURING NONVISIBLE LIGHT: NONOPTICAL ASTRONOMY

Looking back at Figure 3-7, you can see that visible light represents a very tiny fraction of the electromagnetic spectrum. As late as the 1940s, astronomers had no idea how much nonvisible radiation is emitted by objects in space. However, today we know that many objects in space emit undetectable amounts of visible light but release large amounts of nonvisible radiation.

3-11 Specially designed telescopes gather electromagnetic energy in all of the nonvisible parts of the spectrum

From the radio waves and infrared radiation being emitted by vast interstellar gas clouds to the ultraviolet radiation and X-rays from the remnants of stars to bursts of gamma rays of extraordinary power from merging black holes and other sources, our growing ability to see the entire electromagnetic spectrum is revealing intriguing

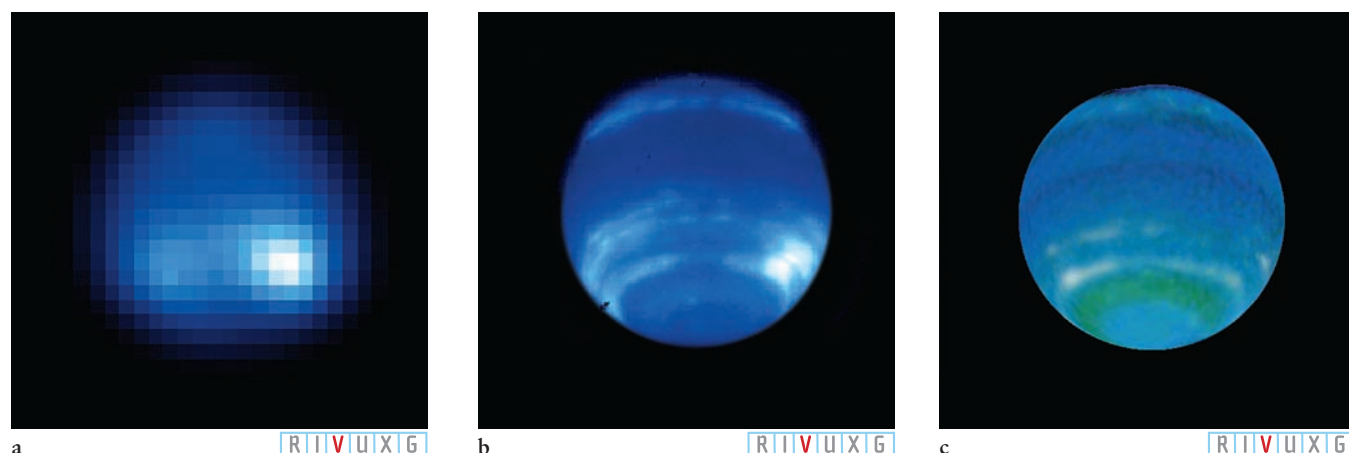


FIGURE 3-30 Images from Earth and Space (a) Image of Neptune from an Earth-based telescope without adaptive optics. (b) Image of Neptune from the same Earth-based telescope with adaptive optics. (c) Image of Neptune from the

Hubble Space Telescope, which does not incorporate adaptive optics technology. (a & b: Courtesy of Center for Adaptive Optics, University of California; c: NASA, L. Sromovksy, and P. Fry, University of Wisconsin-Madison)



RTVUXIG

FIGURE 3-31 The 10-m Keck Telescopes Located on the dormant (and hopefully extinct) Mauna Kea volcano in Hawaii, these huge twin telescopes each consist of 36 hexagonal mirrors measuring 1.8 m (5.9 ft) across. Each Keck telescope has the light-gathering, resolving, and magnifying ability of a

single mirror 10 m in diameter. **Inset:** View down the Keck I telescope. The hexagonal apparatus near the top of the photograph shows the housing for the 1.4-m secondary mirror. (Enrico Sacchetti/Science Source; inset: Laurie Hatch Photography)

phenomena. To get a sense of how much we do not see in the visible part of the spectrum, look at **Figure 3-32**, which shows visible, ultraviolet, and infrared images of the familiar constellation Orion.

A radio telescope uses a large concave dish to collect radio waves The first evidence of nonvisible radiation from outer space came from the work of a young radio engineer, Karl Jansky, at Bell Laboratories. Using radio antennas, Jansky was investigating the sources of static that affect short-wavelength radiotelephone communication. In 1932, he realized that a certain kind of radio noise is strongest when the constellation Sagittarius is high in the sky. Because the center of our Milky Way Galaxy is located in the direction of Sagittarius, Jansky correctly concluded that he was detecting radio waves from elsewhere in the Galaxy (our galaxy is capitalized throughout the book). Jansky's accidental discovery of radio signals from space led to a worldwide effort to

Insight Into Science

Think “Outside the Box” Observations and experiments often require making connections between seemingly unrelated concepts. Jansky's proposal that some radio waves originate in space is an example of connecting apparently disparate fields—astronomy and radio engineering.

“see” what nonvisible wavelengths could teach us about the cosmos.

Radio telescopes record radio signals from the sky. Just as a mirror reflects visible light, the metal surfaces of radio telescopes reflect radio waves. Each radio telescope has a large concave dish (**Figure 3-33**) that focuses radio photons in the same way that an optical telescope

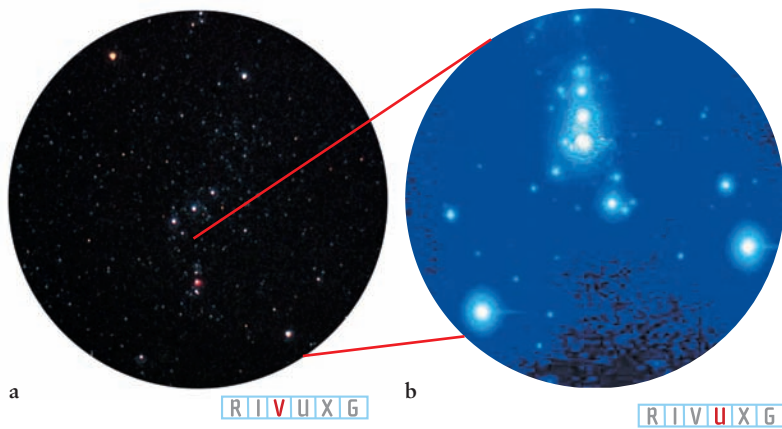


FIGURE 3-32 Orion as Seen in Visible, Ultraviolet, and Infrared Wavelengths (a) An ordinary optical photograph of the constellation Orion. (b) An ultraviolet image of Orion. (c) A false-color view from the Infrared Astronomical Satellite of the entire Orion asterism. Different colors indicate different intensities of infrared radiation. Clearly, different wavelengths provide different information about various objects in space. (a: iStockphoto/Thinkstock; b: NASA/MSX/Johns Hopkins University Applied Physics Laboratory; c: NASA/JPL-Caltech)

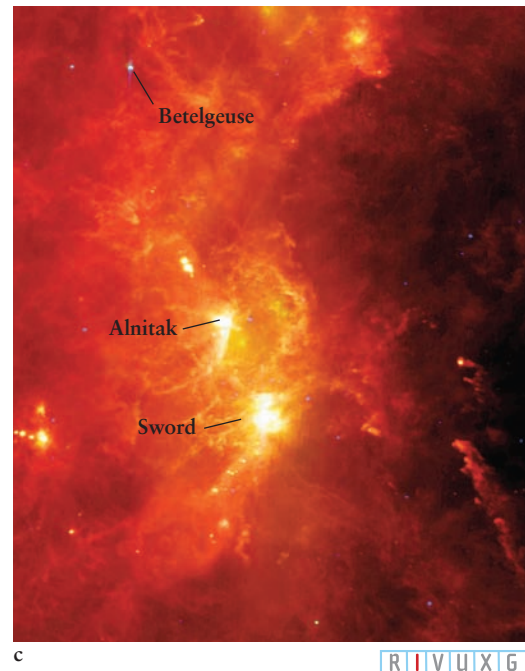


FIGURE 3-33 A Radio Telescope Recall that the secondary mirror or prime focus on most telescopes blocks incoming light or other radiation. This radio telescope at the National Radio Astronomy Observatory in Green Bank, West Virginia, has its prime focus hardware located off-center from the telescope's 100-m by 110-m oval reflector. By using this new design, there is no such loss of signal from equipment covering part of the telescope, unlike the effects of secondary mirrors on optical telescopes. Such configurations are also common on microwave dishes used to receive satellite transmissions for home televisions. (NRAO/AUI/NSF)

mirror focuses visible photons. A small antenna tuned to the desired wavelength is located at the telescope's prime focus or Cassegrain focus. The incoming signal is relayed to amplifiers and recording instruments.

Radio waves have the longest wavelengths of all electromagnetic radiation. Because the angular resolution of any telescope decreases as the wavelengths it collects increases, a radio telescope gives a fuzzier picture than any other type of telescope with the same diameter. Indeed, the first small radio telescopes produced blurry, indistinct images.

Very large radio telescopes create sharper radio images because, as with optical telescopes, the larger the radio dish's diameter, the higher its angular resolution. For this reason, most modern radio telescopes have reflectors more than 25 m in diameter. Nevertheless, even the largest radio dish in existence (305-m-diameter in Arecibo, Puerto Rico) cannot come close to the resolution of the best optical telescopes. For example, a 6-m optical telescope has 2000 times better resolution than a 6-m radio telescope that detects radio waves of 1-mm wavelength.

To overcome the limitation on resolution set by telescope diameter, radio astronomers often use interferometry to produce high-resolution radio images. Recall from Section 3-10 that interferometry combines the data received simultaneously by two or more telescopes. The telescopes can be kilometers, continents, or even worlds apart. The radio signals received by all of the dishes are made to “interfere,” or blend together, and, with suitable computer-aided processing, the combined image of the source is sharp and clear. The results are impressive: The resolution of such a system is equivalent to that of one gigantic dish with a diameter equal to the distance between the farthest telescopes in the array.

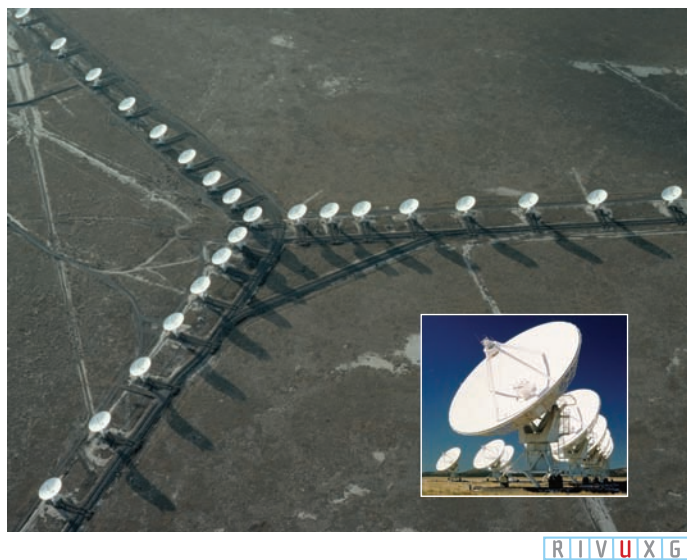


FIGURE 3-34 The Karl G. Jansky Very Large Array The 27 radio telescopes of the Jansky VLA system are arranged along the arms of a Y in central New Mexico. Besides being able to change the angles at which they observe the sky, astronomers can also move these telescopes by train cars so that the array can detect either wide areas of the sky (when they are close together, as in this photograph) or small areas with higher resolution (when they are farther apart). The inset shows the traditional secondary mirror assembly in the center of each of these antennas. (Jim Sugar/Corbis; inset: David Nunuk/Science Source)

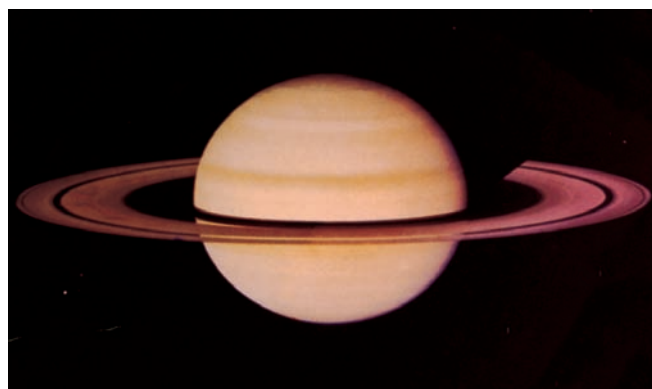
Interferometry, exploited for the first time in the late 1940s, gave astronomers their first detailed views of “radio objects” in the sky. One of the most complex systems of radio telescopes used in this way began operating in 1980 on the plains of San Agustin near Socorro, New Mexico. Originally called the Very Large Array (VLA) and now upgraded and renamed the Karl G. Jansky Very Large Array, this system consists of 27 concave dishes, each 26 m (85 ft) in diameter (**Figure 3-34**). The

27 telescopes, which can be moved, are positioned along the three arms of a gigantic Y that can span a distance of 36 km (22 mi). Working together, they can create radio images with 0.1” resolution. This system produces radio views of the sky with resolutions comparable to those of the best optical telescopes.

More recently, radio telescopes separated by thousands of kilometers have been linked together in **very-long-baseline interferometry (VLBI)**. The Very Long Baseline Array (VLBA) consists of ten 25-m radio telescopes located across the United States, from Hawaii to New Hampshire. With a maximum baseline of 8000 km, VLBA has a resolving power of 0.001”. The best angular resolution on Earth is obtained by combining radio data from telescopes on opposite sides of our planet. In that case, features as small as 0.00001” can be distinguished at radio wavelengths—10,000 times sharper than the best views obtainable from single optical telescopes. Radio telescopes are also being put into space and used in even longer-baseline interferometers.

To make radio images more comprehensible, radio astronomers often use gray scales (see **Figure 1-30i**) or false colors (**Figure 3-35**). The most intense radio emission is shown in red, the least intense in blue. Intermediate colors of the rainbow represent intermediate levels of radio intensity. Black indicates that there is no detectable radio radiation. Astronomers who work at other nonvisible-wavelength ranges also frequently use false-color techniques to display views obtained from their instruments.

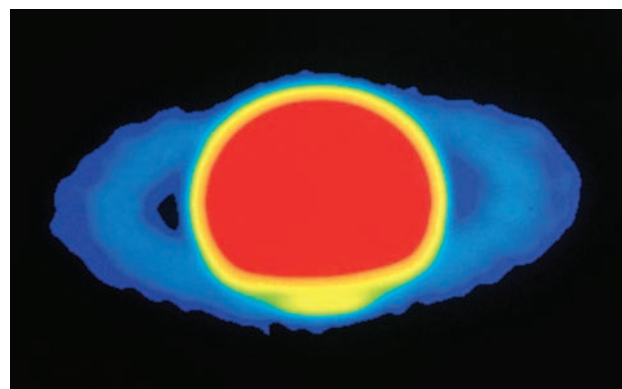
On the boundary between radio and infrared radiation, several microwave telescopes have been orbited around Earth. Notable among these are the Cosmic Background Explorer (COBE), launched in 1989, the Wilkinson Microwave Anisotropy Probe (WMAP), launched in 2001 (both no longer operational), and Odin (2001) and Planck (2009), both of which are still in use. Among other things, these telescopes observe



a

R I V I U X G

FIGURE 3-35 Visible and Radio Views of Saturn (a) This picture was taken by a camera on board a spacecraft as it approached Saturn. The view was produced by sunlight scattered from the planet’s cloudtops and rings. (b) This false-color



b

R I V I U X G

picture, taken by the VLA (the predecessor of the Jansky VLA), shows radio emission from Saturn at a wavelength of 2 cm. (a: NASA; b: Image courtesy of NRAO/AUI/NSF)

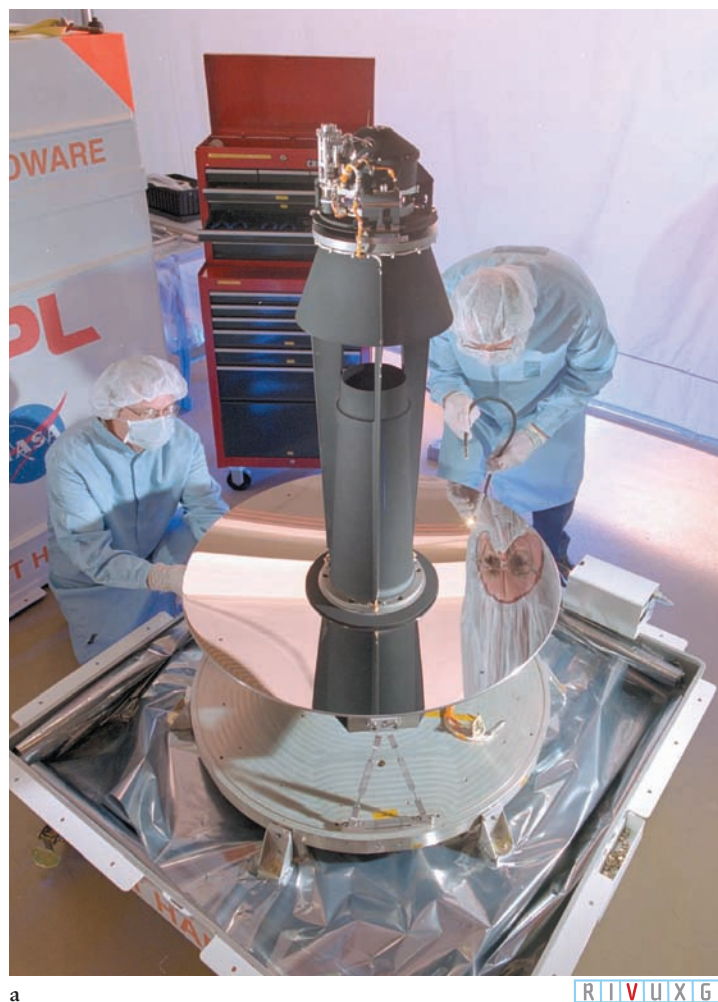
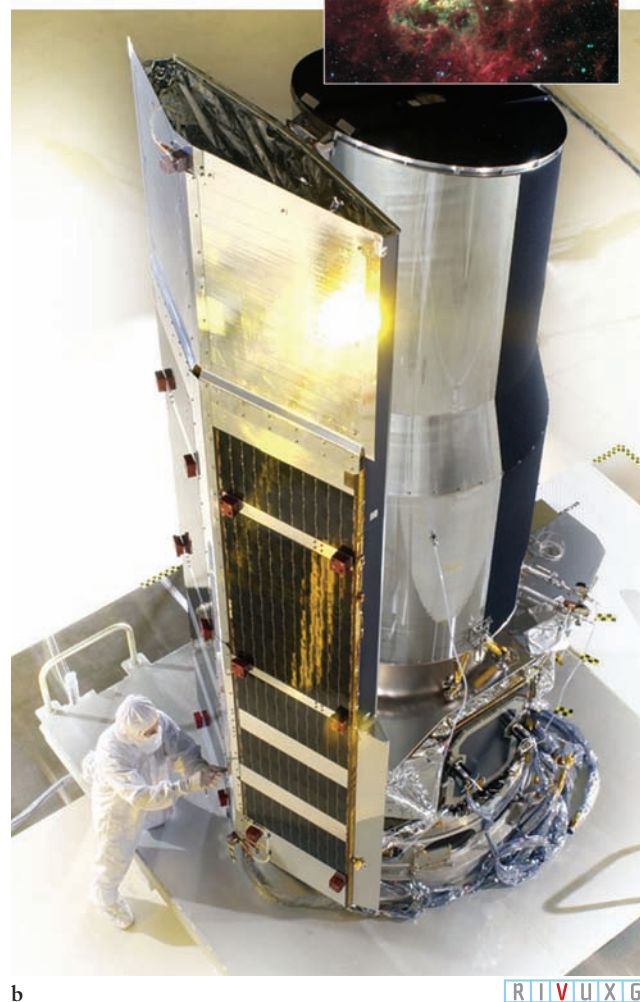


FIGURE 3-36 Spitzer Space Telescope (a) The mirror assembly for the Spitzer Space Telescope showing the 85-cm objective mirror. (b) Launched in 2003, this Great Observatory is taking images and spectra of planets, comets, gas, and dust around other stars



and in interstellar space, galaxies, and the large-scale distribution of matter in the universe. **Inset:** An infrared image of a region of star formation invisible to optical telescopes. (a: Ball Aerospace/NASA/JPL-Caltech; b: NASA/JPL-Caltech; inset: NASA/JPL)

star formation, atoms of different kinds in space, and perhaps most importantly, they gather data that provide us with the temperature of the universe in different directions, results that we will discuss in Chapter 14.

Infrared and ultraviolet telescopes also use reflectors to collect their electromagnetic radiation As with radio and microwave telescopes, infrared and ultraviolet telescopes are all reflectors. Indeed, using suitable CCDs, optical reflecting telescopes can also detect infrared and ultraviolet photons with wavelengths near visible light. The most sensitive infrared-detecting CCDs must be cooled to prevent the heat (and, hence, infrared photons) of the telescope from overwhelming the infrared radiation received from objects in space. Because water vapor is the main absorber of infrared radiation from space,

locating infrared observatories at sites of low humidity can overcome much of the atmosphere's hindrance. For example, the summit of Mauna Kea is exceptionally dry (most of the moisture in the air is below the height of this volcano). Infrared observations are made on the summit by the primarily optical-wavelength Subaru and two Keck telescopes, as well as by NASA's 3-m Infrared Telescope Facility (IRTF).

The best way to avoid the obscuring effects of water vapor is to place a telescope in orbit around Earth. The 1983 Infrared Astronomical Satellite (IRAS) and the Infrared Space Observatory (ISO) launched in 1995 (both now out of service), and Hubble Space Telescope's Near Infrared Camera and Multi-Object Spectrometer (NICMOS) and the Spitzer Space Telescope launched in 2003 (Figure 3-36), among others, have done much to reveal

the full richness and variety of the infrared sky. Spitzer is the infrared equivalent to HST, one of NASA's Great Observatories. In 2010, NASA began flying a telescope called SOFIA (Stratospheric Observatory for Infrared Astronomy) in a modified Boeing 747. At altitude, a giant door in the side of the aircraft slides open and the telescope observes the night sky high above water vapor in the atmosphere. Perhaps the most exciting recent development in infrared astronomy is design and ongoing construction of the 6.5-m-diameter James Webb Space Telescope. This new telescope is scheduled for launch in 2018.

Focus Question 3-9

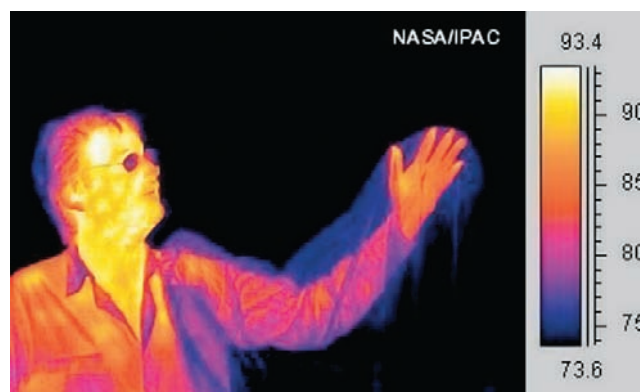
Where in a typical house would infrared detectors indicate the most activity?

Infrared telescopes detect radiation from small bodies in the solar system, as well as from the clouds of dust in our Galaxy, newly formed stars, the dust disks around stars at various stages of their evolution, and the most distant galaxies, whose radiation arrives here primarily at infrared wavelengths, among other infrared-emitting objects. Astronomers have located more than a half-million infrared sources in the sky, most of which are invisible at optical wavelengths. As with radio waves, some infrared radiation passes through interstellar clouds of dust and gas without being scattered, allowing astronomers to see, for example, individual stars at the center of our Galaxy some 245 trillion km (153 trillion mi) away (Figure 3-37).



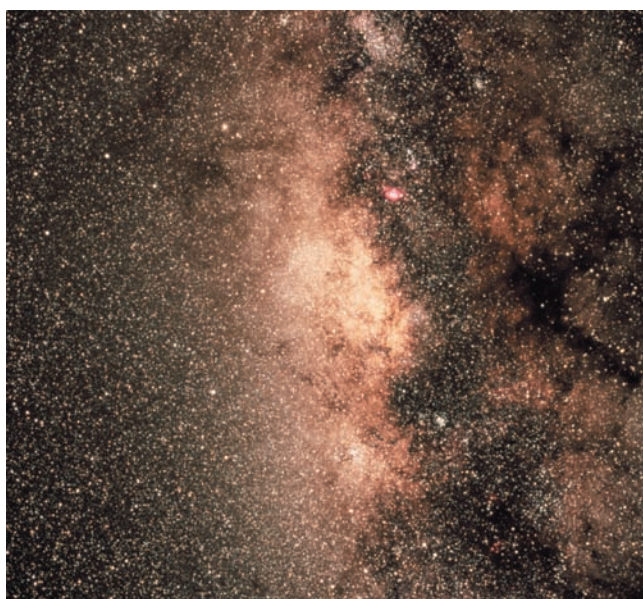
a

R I V U X G



b

R I V U X G



c

R I V U X G



d

R I V U X G

FIGURE 3-37 Views of the Milky Way's Central Regions (a) and (b) Infrared radiation can travel through media that block visible light. (c) An optical image in the direction of Sagittarius, toward the Milky Way's center, showing many nearby stars. The dark regions are interstellar gas and dust clouds that pre-

vent visible light from beyond them from reaching us. (d) An infrared image of the same area of the sky, showing many more distant stars whose infrared radiation passes through the clouds and is collected by our telescopes. (a & b: NASA/JPL-Caltech; c: Dennis di Cicco/CORBIS; d: Science Source)

During the early 1970s, ultraviolet astronomy got off the ground—literally and figuratively. Both *Apollo* and *Skylab* astronauts used small telescopes above Earth's atmosphere to give us some of our first views of the ultraviolet sources in space. Small rockets have also been used to place ultraviolet telescopes briefly above Earth's atmosphere. A typical ultraviolet view is shown in Figure 3-32b.

Some of the finest early ultraviolet astronomy was accomplished by the International Ultraviolet Explorer (IUE), which was launched in 1978 and functioned until 1996. The space shuttles were transformed into orbiting observatories twice in the 1990s, carrying aloft and then returning to Earth three ultraviolet telescopes. In 1992, the Extreme Ultraviolet Explorer (EUVE) was launched. As with infrared observations, ultraviolet images reveal sights previously invisible and often unexpected. Many objects in space emit ultraviolet radiation that astronomers can use to study the chemistries of these cosmic bodies. Therefore, in 1999, astronomers launched the Far Ultraviolet Spectroscopic Explorer (FUSE), which worked through 2007, and the Galactic Evolution Explorer (GALEX), which is in operation today. These latter two telescopes provided information about such things as how much deuterium (hydrogen nuclei with one neutron) was created when the universe formed, the location of particularly hot interstellar gas and dust, and the stellar and chemical evolution of galaxies.

Telescopes dedicated to studying the Sun have observed it from the ground since 1941, and solar telescopes have studied it continuously from space since 1984. Since 1995, the Solar and Heliospheric Observatory (SOHO) has investigated the Sun from space with ultraviolet detectors. More recent orbiting observatories include STEREO (Solar TERrestrial RELations Observatory), launched in 2006, and SDO (the Solar Dynamic Observatory), launched in 2010. These observatories observe the Sun's outer layers rising and sinking, along with a wide range of energetic activity emanating from it. These telescopes show us breathtaking ultraviolet images and movies, some in 3D, of our star. In Chapter 9 we will discuss these observations and their physical origins.

X-ray and gamma-ray telescopes cannot use normal reflectors to gather information X-rays and gamma rays from space interact strongly with the particles in Earth's atmosphere, preventing these dangerous radiations from reaching our planet's surface. Therefore, direct observations of astronomical sources that emit these extremely short wavelengths must be made from space. Astronomers got their first look at X-ray sources during brief rocket flights in the late 1940s. Several small satellites, launched during the early 1970s, viewed the sources of both X-rays and gamma rays in space, revealing hundreds of previously unknown objects, including sev-

eral black holes (see Chapter 12). X-ray telescopes have also been carried on space shuttle missions. The insets in **Figure 3-38** show how different our Sun appears when seen through X-ray and visible-light “eyes.”

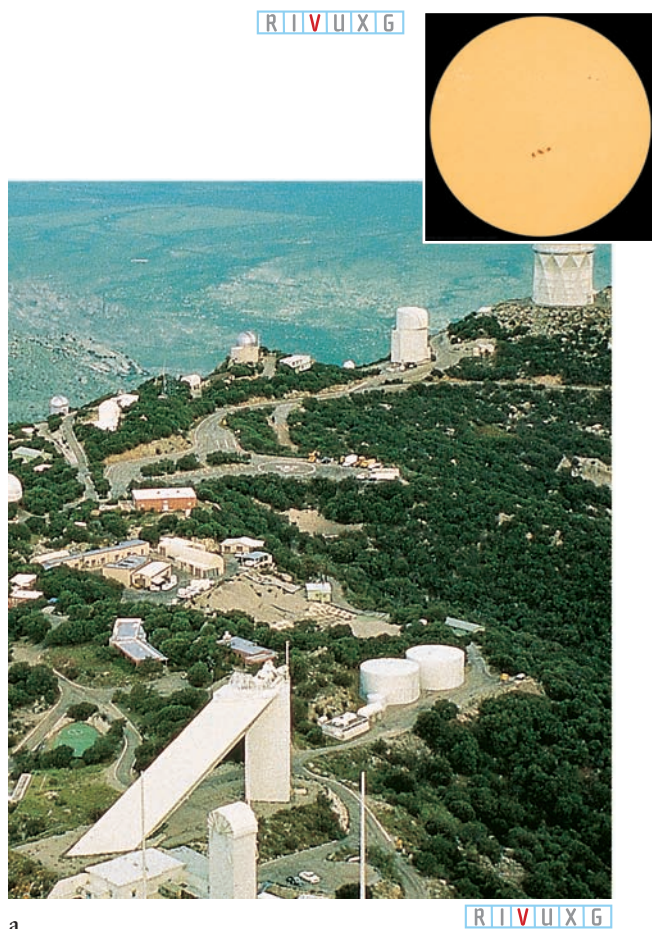
X-ray photons are tricky to collect. Because of their high energies, X-rays penetrate even highly polished surfaces that they meet nearly head-on. Therefore, normal reflecting mirrors cannot be used to focus them. Instead, X-ray telescopes are designed to deflect photons at a fairly shallow angle because X-rays that are barely skimming or grazing a surface can be reflected and thereby focused. This process is analogous to skipping a flat rock on water; throw it at a steep angle and it immediately sinks; throw it at a shallow angle and it will skip off the surface. **Figure 3-39** shows the design of such “grazing incidence” X-ray telescopes.

After being focused, X-rays are detected in several ways. As at infrared, visible, and ultraviolet wavelengths, CCDs can detect these photons. Other devices, called *scintillators*, detect the visible light created as X-rays pass through them. Still other instruments, called *calorimeters*, detect the heat generated by X-rays passing through the detector.

At least seven X-ray telescopes are orbiting Earth today. Among them, the Chandra X-ray Observatory, a NASA Great Observatory named after the Nobel laureate astronomer Subrahmanyan Chandrasekhar, provides images with better than 1" resolution, and the European Space Agency's XMM-Newton has 6" resolution. Other X-ray telescopes are carried high into the atmosphere by balloons. The balloons float at altitudes up to 40 km (25 mi) above Earth's surface and often stay up for several weeks. Whereas low-energy X-rays do not penetrate that far into the atmosphere, higher-energy X-rays do, giving these balloon-borne telescopes a wide range of objects to study.

More than 10,000 X-ray sources have been discovered all across the sky. Among these are planetary atmospheres, stars (see the Sun in Figure 3-38), stellar remnants, vast clouds of intergalactic gas, jets of gas emitted by galaxies, black holes, quasars, clusters of galaxies, and a diffuse X-ray glow that fills the universe.

The electromagnetic radiation with the shortest wavelengths and the most energy are gamma rays. In 1991, the Compton Gamma-Ray Observatory, also a NASA Great Observatory, was carried into orbit by the space shuttle. Named in honor of Arthur Holly Compton, an American physicist and Nobel laureate who made important discoveries about gamma rays, this orbiting observatory carried four instruments that performed a variety of observations, giving us tantalizing views of the gamma-ray sky until the telescope failed in 2000. At least four gamma-ray telescopes, including the *SWIFT* Gamma-Ray Burst Mission and the Fermi Large Area Telescope, are presently operating in space, and several others have been flown on balloon missions.



a



b

FIGURE 3-38 Nonvisible and Visible Radiation (a) The McMath-Pierce Solar Telescope at Kitt Peak Observatory near Tucson, Arizona (the inverted V-shaped structure), takes visible-light photographs of the Sun, such as the one shown in the inset. (b) This X-ray telescope was carried aloft in 1994 by a space shuttle. The inset shows an X-ray image of the Sun. Comparing the images in the two insets reveals how important

observing nonvisible radiation from astronomical phenomena is to furthering our understanding of how the universe operates. (a: NOAO/AURA/NSF; inset: NASA/SDO/HMI; b: NASA; inset: Solar X-ray image from the Yohkoh mission of ISAS, Japan; the X-ray telescope was prepared by the Lockheed-Martin Solar and Astrophysics Laboratory, the National Astronomical Observatory of Japan, and the University of Tokyo with the support of NASA and ISAS)

Gamma-ray telescopes, such as the VERITAS and MAGIC arrays, are also operating here on Earth. As we noted earlier, gamma rays do not reach Earth's surface from space. However, when these photons strike atoms in Earth's atmosphere, they cause cascades of gas particles to rush earthward at high velocities. As these particles descend into the atmosphere, they give off visible and UV radiation called *Cherenkov radiation*. This radiation travels in predictable angles from the particles that emit it and therefore detecting it enables astronomers to calculate from what direction the original gamma ray came.

Gamma rays are too powerful even for grazing incidence telescopes. Therefore, astronomers have devised other methods of detecting them and determining their origin. These include absorbing them in crystals; allowing them to pass through tiny holes called *collimators*

whose directions are well determined; and using chambers in which the gamma rays transform into electrons and positrons (positively charged electrons), leaving a track whose direction can be determined. These techniques are not nearly as precise as those used in other parts of the spectrum, and the best-resolution gamma-ray instruments are only accurate to about 5'. Several thousand gamma-ray sources have been discovered.

We now have telescopes with which we can see the energy in the universe from virtually all parts of the electromagnetic spectrum (Figure 3-40). Telescopes provide us with more than just stunning images of objects in space. They also provide information about the chemistry of stars and interstellar gas and dust, the motion of stars and galaxies toward or away from us, whether stars are rotating, the formation of galaxies and their groupings in the universe, and the expansion of the uni-

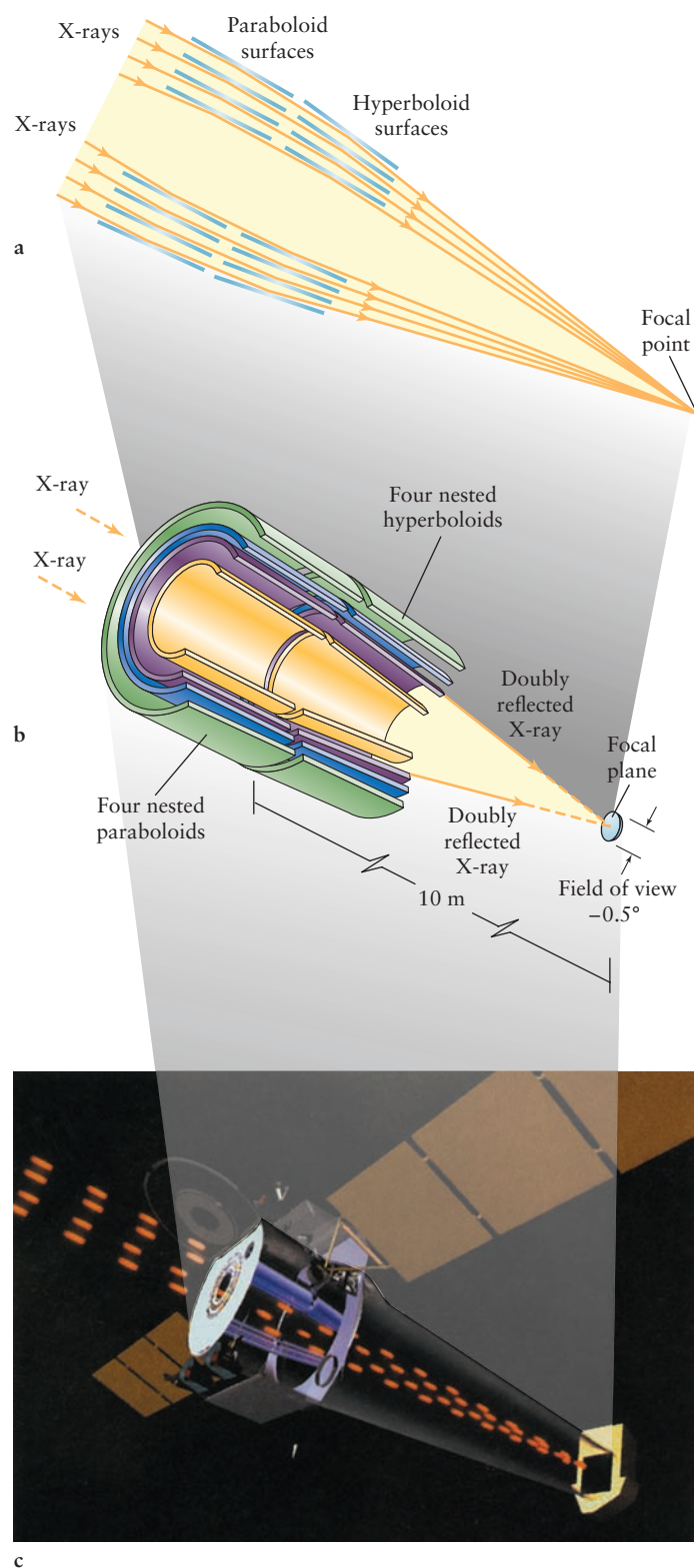


FIGURE 3-39 Grazing Incidence X-ray Telescopes (a) X-rays penetrate objects they strike head-on. In order to be focused, X-rays have to be gently nudged by skimming off cylindrical “mirrors.” (b) The shapes of the mirrors optimize the focus. (c) The diagram shows how X-rays are focused in the Chandra X-ray Telescope. (a & b: NASA/JPL-Caltech; c: NASA/CXC/D. Berry & A. Hobart)

verse and its fate, among myriad other things. We will explore how we get this information when we study the various objects in space.

So far in this chapter, we have explored the basic properties of electromagnetic radiation and the techniques astronomers use to collect it. To understand electromagnetic radiation more completely, we now delve more deeply into the properties of matter. By the end of this chapter, you will have learned how and why different kinds of matter emit and absorb different wavelengths of electromagnetic radiation and other related properties of matter and energy.

BLACKBODY RADIATION

We begin our study of why we see what we see with a familiar property of matter—temperature. We will discover that for many astronomical objects, their temperatures determine the relative numbers of photons that they emit at all different wavelengths.

3-12 An object’s peak color shifts to shorter wavelengths as it is heated

Imagine that you have mounted an iron rod in a vise in a completely darkened room. You cannot see the rod because it emits too few visible photons, and you cannot feel its warmth without touching it because it is at the same temperature as the air in the room. Now imagine that there is also a propane torch in the room. You light it and heat the iron rod for several seconds, until it begins to glow. At the same time, you can feel the rod warm up, meaning that it is also emitting more infrared radiation.

6 The rod’s first visible color is red—it glows “red hot” (Figure 3-41a). Heated a little more, the rod appears orange and brighter and feels hotter (Figure 3-41b). Heated more, it appears yellow and brighter still and feels even hotter (Figure 3-41c). After even more heating, the rod appears white-hot and brighter yet. If it does not melt, further heating will make the rod appear blue and even brighter and feel even hotter. As confirmed by the photographs in Figure 3-41, this experiment shows how the amount of electromagnetic radiation that any object emits changes with its temperature. A “red-hot” object is the coolest of all glowing bodies.

If you were to put the light emitted by a heated object, like the iron rod, through a prism or diffraction grating (which separates the colors of light, just like a prism), you would discover that all wavelengths are present and that one is brightest. This peak wavelength is the color that the object appears to have. The experiment with the rod reveals two principles about the electromagnetic radiation emitted by objects as their temperatures change:

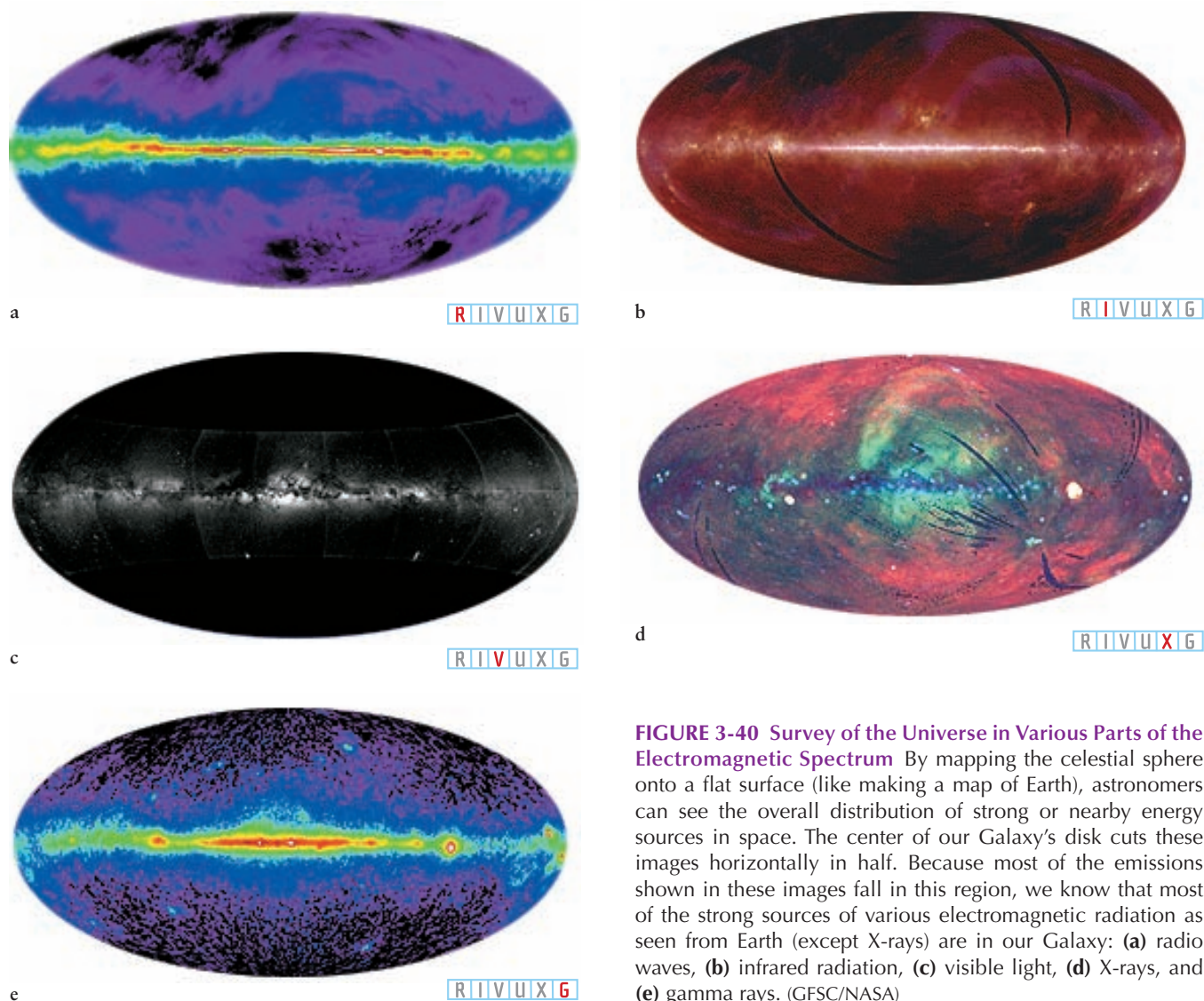


FIGURE 3-40 Survey of the Universe in Various Parts of the Electromagnetic Spectrum By mapping the celestial sphere onto a flat surface (like making a map of Earth), astronomers can see the overall distribution of strong or nearby energy sources in space. The center of our Galaxy's disk cuts these images horizontally in half. Because most of the emissions shown in these images fall in this region, we know that most of the strong sources of various electromagnetic radiation as seen from Earth (except X-rays) are in our Galaxy: **(a)** radio waves, **(b)** infrared radiation, **(c)** visible light, **(d)** X-rays, and **(e)** gamma rays. (GFSC/NASA)

1. As an object heats up, it gets brighter by emitting more electromagnetic radiation at all wavelengths.
2. The brightest color (most intense wavelength) of the emitted radiation changes with temperature.

Most objects big enough to be seen with the naked eye continuously emit electromagnetic radiation as just described. The peak wavelength of any object's radiation, that is, where it emits most intensely, is denoted λ_{max} . When the object is relatively cool, such as a rock or an animal, λ_{max} is a radio, microwave, or infrared wavelength. When it is hot enough, like a fire or the Sun, λ_{max} is in the range of visible light, giving a hot object its characteristic color. Exceptionally hot objects, such as some stars, emit λ_{max} in the ultraviolet part of the electromagnetic spectrum.

Stars, molten rock, and iron bars are approximations of an important class of objects that scientists call **blackbodies**. An ideal blackbody absorbs all of the electromagnetic radiation that strikes it. This incoming radiation heats up the blackbody, which then reemits the energy it has absorbed, but with different intensities at each wavelength than it received. Furthermore, the pattern of radiation emitted by blackbodies is independent of their chemical compositions. By measuring the intensity of radiation emitted by a blackbody at several wavelengths, it is possible to plot a curve of its emissions over all wavelengths (**Figure 3-42**). Ideal blackbodies have smooth **blackbody curves**, whereas objects that approximate blackbodies, such as the Sun, have more jagged curves whose variations from the ideal blackbody are caused by other physics (**Figure 3-43**).

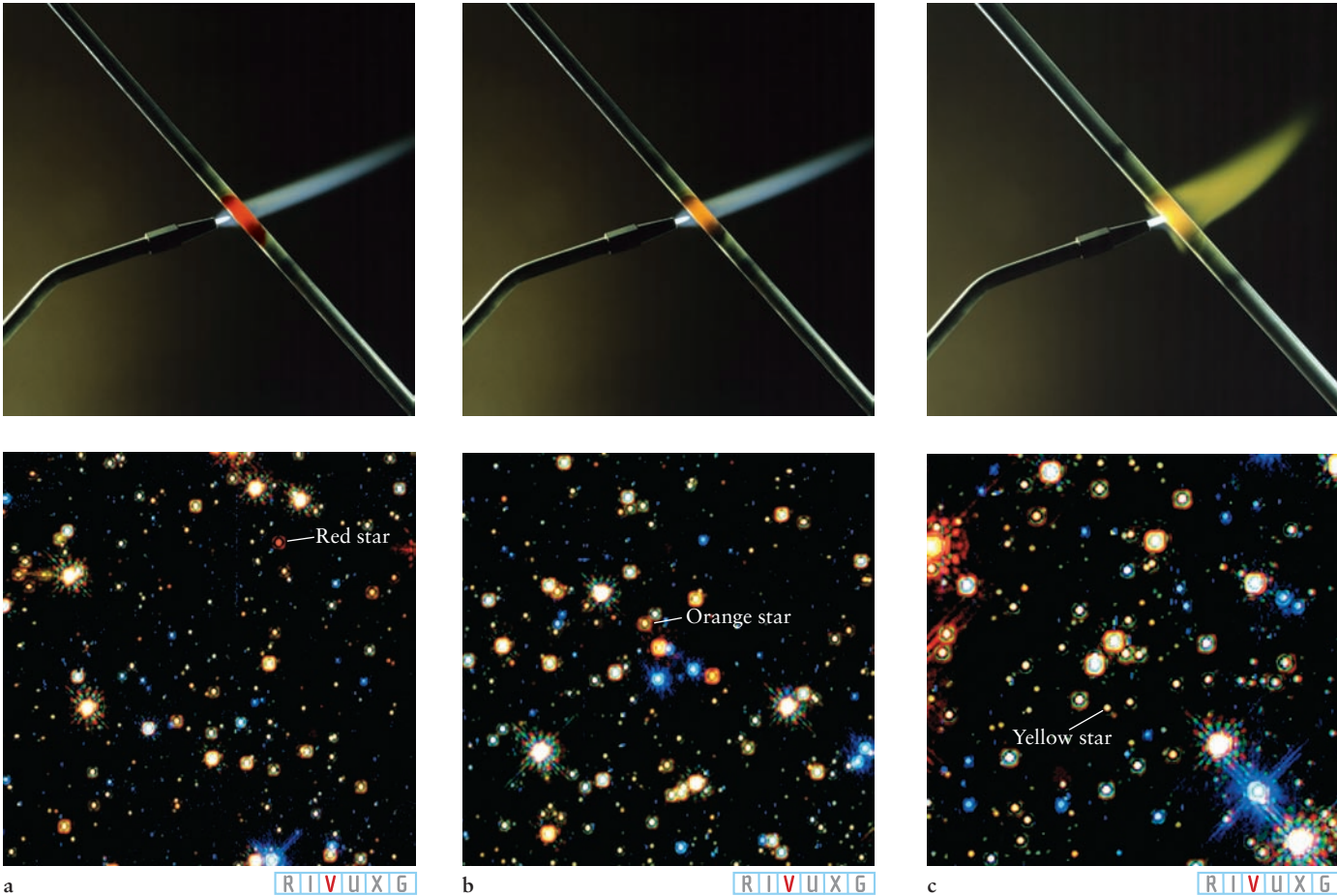


FIGURE 3-41 Objects at Different Temperatures Have Different Colors and Brightnesses This sequence of photographs shows the changing appearance of a piece of iron as it is heated. As the temperature increases, the amount of energy radiated by the bar increases, and so it appears brighter. The apparent color of the bar also changes because, as the temperature increases,

the peak wavelength of light emitted by the bar decreases. The labeled stars shown have roughly the same temperatures as the bars above them. (top photos: © 1984 Richard Megna/Fundamental Photographs; bottom photos: © Don Figer [Space Telescope Science Institute] and NASA)

The total amount of radiation emitted by a blackbody at each wavelength depends only on the object's temperature and how much surface area it has. The bigger an object is, the brighter it is at all wavelengths. However, the *relative* amounts of different wavelengths (for example, the intensity of light at 750 nm compared to the intensity at 425 nm) depend on just the body's temperature. So, by examining the relative intensities of an object's blackbody curve, we are able to determine its temperature, regardless of how big or how far away it is. This process is analogous to how a thermometer tells your internal temperature no matter how big you are. (Temperatures throughout this book are expressed in kelvins. If you are not familiar with the Kelvin temperature scale, review Appendix B: Temperature Scales.)

Rather than absorb and reradiate light from an outside source like our iron rod, stars radiate electro-

magnetic radiation that is generated inside them. Even so, stars behave nearly like blackbodies, and the self-generated radiation they emit follows the idealized blackbody curves fairly well.

Insight Into Science

Many Phenomena Result from Several Effects

Recall from Section 3-1 that white is not a fundamental color. Why then do objects at certain temperatures appear “white-hot,” whereas cooler ones are red, orange, or yellow, and hotter ones appear blue or violet? The answer is that when blackbodies peak in the green part of the spectrum, all the color-sensitive cones in our eyes are stimulated about evenly, an effect that the brain interprets as “white.”

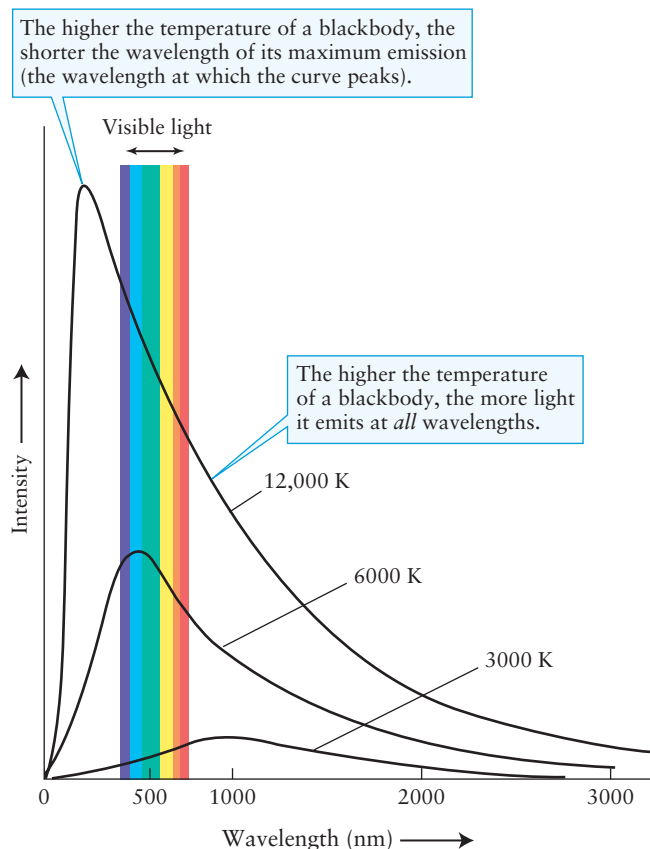


FIGURE 3-42 Blackbody Curves Three representative blackbody curves are shown here. Each curve shows the intensity of radiation (vertical axis) over a wide range of wavelengths (horizontal axis) emitted by a blackbody at a particular temperature.

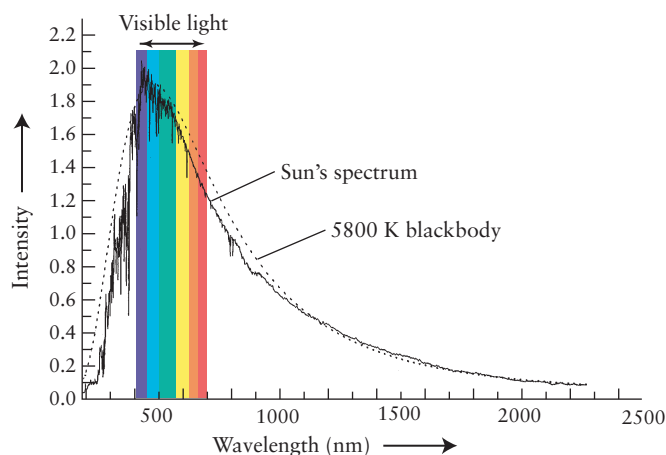


FIGURE 3-43 The Sun as a Blackbody This graph shows the similarity between the intensity of sunlight and the intensity of radiation from a 5800 K blackbody over a wide range of wavelengths. Earth's atmosphere scatters shorter-wavelength photons more than it scatters longer-wavelength ones. As seen from Earth, this scattering lowers the intensity of sunlight and shifts the blackbody peak to the right from where it would occur normally. To avoid these effects, the measurements of the Sun's intensity in this figure were made in space.

3-13 The relative intensities of different emitted colors reveal a star's surface temperature

In 1893, the German physicist and 1911 Nobel laureate Wilhelm Wien showed that

The peak wavelength of radiation emitted by a blackbody is inversely proportional to its temperature.

In other words, the hotter any object becomes, the shorter is its λ_{max} , and vice versa. **Wien's law**, often called *Wien's displacement law*, is the mathematical relationship between the location of the peak for each curve in Figure 3-42 and that blackbody's temperature.

Wien's law proves very useful in computing the surface temperature of a star. It says that to determine that temperature, all we need to find is the peak wavelength of its electromagnetic radiation—we do not need to know the star's size, distance, or any other physical property.

In 1879, the Austrian physicist Josef Stefan observed that

An object emits energy per unit area at a rate proportional to the fourth power of its temperature in kelvins.

Think of the electric heating element on a stove. As this device becomes hotter, it gets brighter and emits more energy. Stefan's result says that if you double the temperature of an object (for example, from 500 to 1000 K), the energy emitted each second from each piece of the object's surface increases by a factor of 2^4 , or 16 times. If you triple the temperature (for example, from 500 to 1500 K), the rate at which energy is emitted increases by a factor of 3^4 , or 81 times. Stefan's results were put on a firm theoretical foundation by Ludwig Boltzmann in 1885. In their honor, the intensity–temperature relationship for blackbodies is named the **Stefan-Boltzmann law**.

The Stefan-Boltzmann law and Wien's law are powerful tools that describe two basic properties of blackbody radiation, namely, the relative rate at which photons of all different wavelengths are emitted and the wavelength that is emitted most intensely.

Blackbody curves, such as those shown in Figure 3-42, provide the same information as Wien's law and the Stefan-Boltzmann law, as well as show the intensity of each blackbody at all wavelengths. Note that the blackbody curves for objects with different temperatures never touch or cross. Therefore, when astronomers measure the intensity of radiation from a star at a few wavelengths, they can determine the object's temperature by fitting these measurements to the appropriate, unique blackbody curve.

Although intensity curves of stars closely “follow” blackbody curves, they are not ideal blackbodies. The

differences between an ideal blackbody curve and the curves seen from actual stars reveal stellar chemistries, the presence of companion stars and planets too dim to see, and the motions of stars toward or away from us, among other information. We begin studying these issues by considering the actual spectrum of the Sun.

7 Figure 3-43 shows how the intensity of sunlight varies with wavelength. The blackbody curve for a body with a temperature of 5800 K is also plotted in this figure. Note how the observed intensity curve for the Sun nearly reproduces the ideal blackbody curve at most wavelengths. The differences in the two curves are due to photons removed from the Sun's blackbody radiation by gases between the Sun's visible surface and our eyes. We discuss this phenomenon further shortly. Because the observed intensity curves for most stars and the idealized blackbody curves are so closely correlated, the laws of blackbody radiation can be applied to starlight. The peak of the intensity curve for the Sun is at a wavelength of about 500 nm, which is in the blue-green part of the visible spectrum.

The shapes of ideal blackbody curves were first derived from basic physical principles (and a deeper understanding of matter and energy than was available to Ludwig Boltzmann) in 1900 by the German physicist Max Planck. To determine these curves, he had to assume that objects receive electromagnetic radiation as individual packets of energy. This surprising assumption was reinforced in 1905 by Albert Einstein, who suggested that light—a wave—also behaves as particles (photons), as discussed in Section 3-3. Planck was awarded the Nobel Prize in 1918 for his ideas about photons. Einstein's 1921 Nobel Prize was for related work (he did not earn it for his development of special or general relativity, which we study in Chapter 12).

We can now describe our experiment with the iron bar at the beginning of this section in terms of the energies and intensities of photons. Recall from Section 3-3 that the energy carried by a photon is inversely proportional to its wavelength. In other words, long-wavelength photons, such as radio waves, carry little energy, while short-wavelength photons, like X-rays and gamma rays, each carry much more energy. When the rod was cool, it mostly emitted lower-energy infrared and radio photons; as it heated up, it emitted more and more photons of all wavelengths and therefore glowed more, eventually emitting mostly visible photons, which is why it looked brighter, felt hotter, and changed color.

The relationship between the intensity of electromagnetic radiation at every wavelength emitted by a given blackbody and its temperature is called **Planck's law** (Table 3-1): Planck's law gives the shapes of the curves in Figure 3-42. To summarize: A cool object emits primarily long-wavelength photons that carry little energy, while a hot object emits mostly short-wavelength photons that carry much more energy. Furthermore, hotter objects emit more photons of all wavelengths than do cooler objects and, according to Wien's law, the hotter a blackbody, the shorter the wavelength of the peak of its emission. In later chapters, we will find these relationships invaluable for understanding how stars of various temperatures interact with gas and dust in space.

Focus Question 3-10

Referring to Figure 3-7 if necessary, where on the horizontal axis are radio waves relative to visible light in Figure 3-42?

Focus Question 3-11

If a blackbody's peak wavelength is in the infrared, the object is often visible to our eyes. What color will it have?

TABLE 3-1 SOME PROPERTIES OF ELECTROMAGNETIC RADIATION

	Wavelength (nm)	Photon energy (eV)*	Peak of blackbody temperature (K)
Radio	$>10^7$	$<10^{-4}$	<0.03
Microwave**	10^7 to 4×10^5	10^{-4} to 3×10^{-3}	0.03 to 30
Infrared	4×10^5 to 7×10^2	3×10^{-3} to 2	30 to 4100
Visible	7×10^2 to 4×10^1	2 to 3	4100 to 7300
Ultraviolet	4×10^2 to 10^1	3 to 10^3	7300 to 3×10^6
X-ray	10^1 to 10^{-2}	10^3 to 10^5	3×10^6 to 3×10^8
Gamma ray	$<10^{-2}$	$>10^5$	$>3 \times 10^8$

Note: $>$ means greater than; $<$ means less than

*1 eV (electron volt) = 1.6×10^{-19} J

**Microwaves, listed here separately, are often classified as radio waves or infrared radiation.

IDENTIFYING THE ELEMENTS BY ANALYZING THEIR UNIQUE SPECTRA

In 1814, the German optician Joseph von Fraunhofer repeated Newton's classic experiment of shining a beam of sunlight through a prism (recall Figure 3-1). However, Fraunhofer magnified the resulting rainbow-colored spectrum. He discovered that the solar spectrum contains hundreds of fine dark lines, which became known as **absorption lines** because the light of these colors has been absorbed by gases between the Sun and the viewer on Earth. Fraunhofer counted more than 600 such lines, and today physicists have detected more than 30,000 of them. Thousands of spectral lines are visible in the photograph of the Sun's spectrum shown in Figure 3-44.

By the mid-1800s, chemists discovered that they could produce spectral lines in the laboratory. Around 1857, the German chemist Robert Bunsen invented a special gas burner that produces a clean, constant flame. Certain chemicals are easy to identify by the distinctive colors emitted when bits of the chemical are sprinkled into the flame of a Bunsen burner.

Bunsen's colleague Gustav Kirchhoff suggested that light from the colored flames could best be studied by passing it through a prism. By separating the colors, the chemists could see exactly which ones were present. The

process of analyzing the spectrum of an object is called *spectroscopy*. Bunsen and Kirchhoff collaborated in designing and constructing the first **spectroscope**. This device consists of a narrow slit, a prism, and several lenses that straighten the light rays and magnify the spectrum so that it can be closely examined (Figure 3-45).

The chemists discovered that the spectrum from a chemical heated by a flame consists of a pattern of thin, bright spectral lines, called **emission lines**, against a dark background (not a blackbody spectrum!). They next found that *the number of lines produced and their colors are unique to the element or compound heated to produce them*. Thus, in 1859, the technique of **spectral analysis**, the identification of chemical substances by their spectral lines, was born.

3-14 Each chemical element produces its own unique set of spectral lines

A chemical **element** is a fundamental substance that cannot be broken down into more basic units while still retaining its properties. By the mid-1800s, chemists had already identified such familiar elements as hydrogen, oxygen, carbon, iron, gold, and silver. Spectral analysis promptly led to the discovery of additional elements, many of which are quite rare.

After Bunsen and Kirchhoff recorded the prominent spectral lines of all of the known elements, they began to

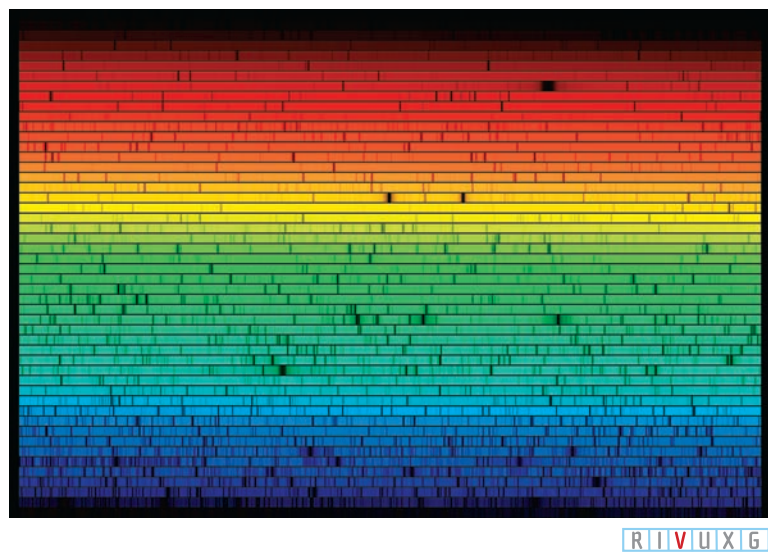


FIGURE 3-44 Solar Spectrum Starlight passing through a prism or diffraction grating spreads into its component colors. From such spectra we can learn an incredible amount about stars, including their masses, surface temperatures, diameters, chemical compositions, rotation rates, and motions toward or away from us. This image, called a *spectrogram*, shows the spectrum of the Sun sliced and stacked to fit on this page. In it you can see thousands of absorption lines. As you study spectra, also note the distinct differences between the intensities of the various colors emitted by each star and by different stars. (N. A. Sharp, NOAO/AURA/NSF)

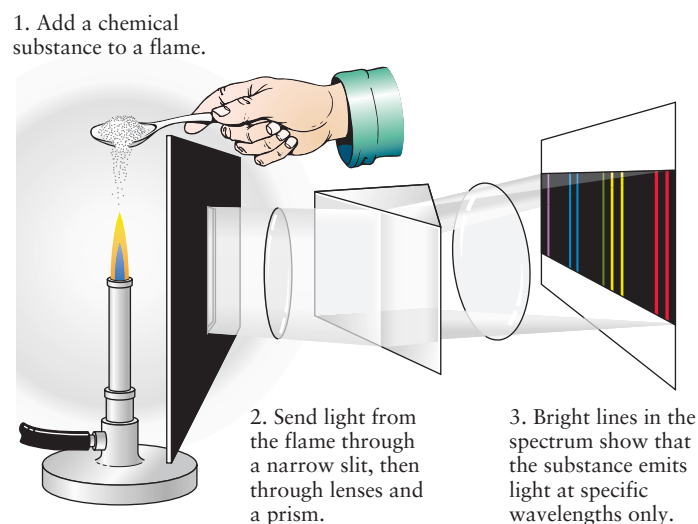


FIGURE 3-45 Early Spectroscope In the mid-1850s, Kirchhoff and Bunsen discovered that when a chemical substance is heated and vaporized, the resulting spectrum exhibits a series of bright spectral lines when passed through a slit, lenses, and a prism. This device is called a spectroscope. In addition, they found that each chemical element produces its own characteristic pattern of spectral lines. The lenses focus and magnify the spectrum.

discover other spectral lines in mineral samples. In 1860, for example, they found a new line in the blue portion of the spectrum of mineral water. After chemically isolating the previously unknown element responsible for the line, they named it *cesium* (from the Latin *caesius*, meaning “gray-blue”). The next year, a new spectral line in the red portion of the spectrum of a mineral sample led to the discovery of the element rubidium (from *rubidus*, for “red”).

During a solar eclipse in 1868, astronomers found a new spectral line in the light coming from the upper atmosphere of the Sun when the main body of the Sun was hidden by the Moon. This line was attributed to a new element, which was named *helium* (from the Greek *helios*, meaning “sun”). Helium was not actually discovered on Earth until 1895, when it was identified in gases obtained from a uranium compound.

The chemical elements are most conveniently displayed in the form of a **periodic table** (Appendix N: Periodic Table of the Elements). In addition to the 92 naturally occurring elements, there are artificially produced ones. All of the human-made elements are heavier than uranium (U), and all are highly radioactive, meaning that they spontaneously decay into lighter elements shortly after being created in the laboratory.

During some experiments, Bunsen and Kirchhoff examined the blackbody spectra of hot sources whose light they passed through cooler gas. Instead of seeing a continuum of wavelengths, as they would have seen had the cool gas not been there, they saw dark *absorption lines* among the colors of the rainbow created by the hot source. (This phenomenon is what Fraunhofer had seen earlier from the Sun—the cooler gases in his results were the atmospheres of Earth and the Sun.) In other experiments, when looking just at gases without hot sources behind them, they saw bright *emission lines* against otherwise dark backgrounds. Combining the results from both sets of experiments, they discovered

that *the emission lines of a particular gas occur at exactly the same wavelengths as the absorption lines of that gas.*

Because each chemical element produces its own unique pattern of spectral lines, scientists can determine the chemical composition of a remote astronomical object by identifying the lines in its spectrum. For example, **Figure 3-46** shows a portion of the Sun’s absorption spectrum in the blue-violet part of the spectrum, along with the emission spectrum of an iron sample taken here on Earth. The spectral lines of iron also appear in the Sun’s spectrum, so we can reasonably conclude that the Sun’s atmosphere contains some vaporized iron. This gaseous iron absorbs and then scatters certain wavelengths from the continuum emitted below it in all directions, preventing that light from reaching Earth.

After photography was invented, scientists preferred to do **spectroscopy** by making a permanent photographic record of spectra. A device for photographing a spectrum is called a spectrograph, and this instrument, in its numerous variations, is the astronomer’s most important tool.

In its earliest form, a spectrograph was attached directly to a telescope. It consisted of a slit, two lenses, and a prism arranged to focus the spectrum of an object on film, similar to the device sketched in Figure 3-45. Although conceptually straightforward, this early type of spectrograph had severe drawbacks. A prism does not spread colors evenly: The blue and violet portions of the spectrum are spread out more than the red portion. In addition, because the blue and violet wavelengths must pass through more glass than the red wavelengths (see Figure 3-1), light is absorbed in the prism unevenly across the spectrum. Indeed, a glass prism is opaque to ultraviolet wavelengths.

Practical spectrographs used in research today (shown conceptually in **Figure 3-47a**) separate light from objects in space into the colors of the rainbow using a **diffraction grating**, which is a piece of glass or

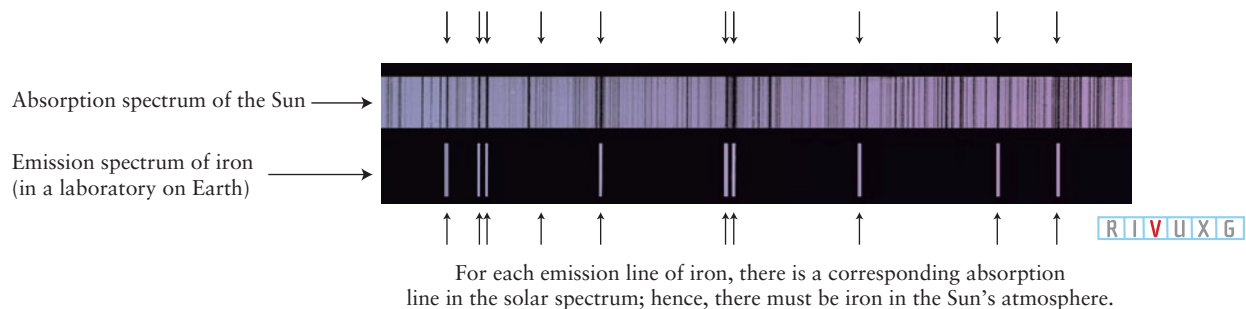


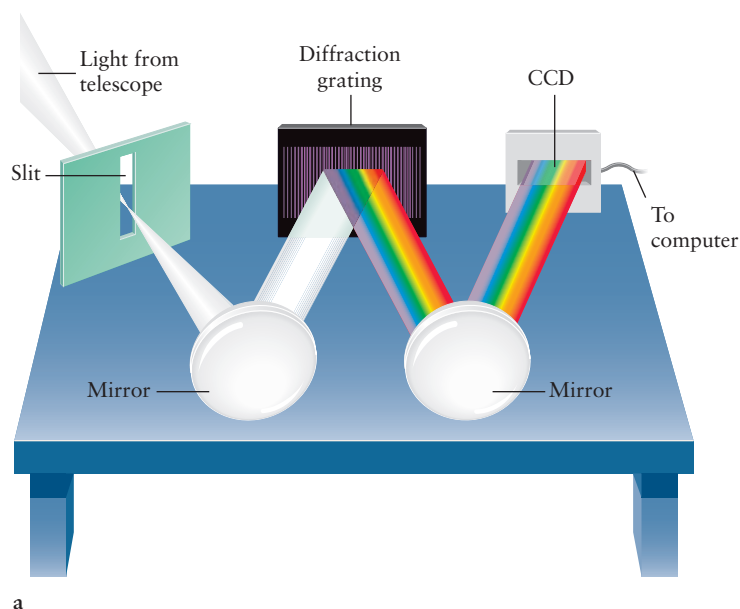
FIGURE 3-46 Iron in the Sun’s Atmosphere The upper (absorption) spectrum is a portion of the Sun’s spectrum from 425 to 430 nm. Numerous dark spectral lines are visible. The lower (emission) spectrum is a corresponding portion of the spectrum

of vaporized iron. Several emission lines can be seen against the black background. The iron lines coincide with some of the solar absorption lines, proving that there is some iron (albeit a very tiny amount) in the Sun’s atmosphere. (Carnegie Observatories)

Insight Into Science

Seek Relationships Finding how properties of different things relate to one another often provides invaluable insights into how they work. The periodic table, for example, enables scientists to determine the properties of similar chemical elements. Keep an eye out for other such relationships throughout this book.

plastic on which thousands of closely spaced parallel grooves are cut or etched. Some of the finest diffraction gratings have more than 10,000 grooves per centimeter. The spacing of the grooves must be very regular. Light waves are diffracted by the grooves in the diffraction grating, just as light or water waves are diffracted when passing through slits (see Figure 3-2). Some diffraction gratings have the light pass through them, while others have the light reflect off their surfaces, as shown



in Figure 3-47a. Different wavelengths are diffracted by different amounts and so they are spread out, just as occurs when the light passes through a glass prism.

Figure 3-47a shows the design of a modern diffraction grating spectrograph with a CCD (charge-coupled device; see Figure 3-25) that has replaced film to record the spectra. This optical device typically mounts at the Nasmyth, coudé, or Cassegrain focal point of a telescope. The image of the object to be examined is focused on the slit. After the spectrum of a star, galaxy, or other object has been recorded, the CCD collects an emission spectrum from a known source, such as a gas composed of helium, neon, and argon, which is also focused on the slit. This “comparison spectrum” is placed next to the spectrum of the object, as in Figure 3-46. Because the wavelengths of the spectral lines in the comparison spectrum are already known from laboratory experiments, these lines can be used to identify and measure the wavelengths of the lines in the spectrum of the star



FIGURE 3-47 A Grating Spectrograph (a) The diffraction grating in a grating spectrograph has many parallel lines on its surface that reflect light of different colors in different directions. This separation of colors (wavelengths) allows the object's spectrum to be analyzed. (b) This peacock feather contains numerous natural diffraction gratings. The role of the parallel lines etched in a human-made diffraction grating is played by parallel rods of the protein melanin in the feathers. (c) CDs and DVDs store information on closely spaced bumps located on a set of nearly parallel tracks. Light striking these tracks systematically reflects different colors in different directions—a CD or a DVD behaves like a diffraction grating. (b: © 1990 Paul Silverman/Fundamental Photos; c: Don Farrall/Photo-disc/Getty Images)

or galaxy under study. Figures 3-47b and 3-47c show diffraction spectra occurring on Earth.

The spectral data are converted by computer to a graph that plots light intensity against wavelength. Dark lines in the rainbow-colored spectrum appear as depressions or valleys on the graph, while bright lines in the spectrum appear as peaks. For example, **Figure 3-48** shows both an absorption spectrum and an emission spectrum for hydrogen. The five absorption lines (Figures 3-48a, b) occur at precisely the same wavelengths as the five emission lines for hydrogen (Figures 3-48c, d).

3-15 The various brightness levels of spectral lines depend on conditions in the spectrum's source

By the early 1860s, Kirchhoff had discovered the conditions under which different types of spectra are observed. His description is summarized today as **Kirchhoff's laws**:

Law 1 *A solid, liquid, or dense gas produces a **continuous spectrum** (also called a **continuum**)—a complete rainbow of colors without any spectral lines. This is a **blackbody spectrum**.*

The light emitted by a hot iron rod and by an incandescent lightbulb are examples of continuous spectra that are bright enough for us to see.

Law 2 *A rarefied (opposite of dense) gas produces an **emission line spectrum**—a series of bright spectral lines against a dark background.*

The light emitted by neon lights and by low-pressure sodium vapor lights are examples of emission line spectra. The neon has bright-red emission lines, while the low-pressure sodium has bright-yellow lines.

Law 3 *The light from an object with a continuous spectrum that passes through a cool gas produces an **absorption line spectrum**—a series of dark spectral lines among the colors of the rainbow.*

Sunlight and the light from other stars pass through several cooler gases on their way to us. Hence, light from stars produces absorption line spectra.

In summary (**Figure 3-49**), a continuum is seen if there is no gas between a blackbody and the observer; absorption lines are seen if the background object is hotter than the gas between the observer and the object; and emission lines are seen if the background is cooler than such a gas.

Consider, for example, the spectrum of the Sun. We know that the Sun's surface emits a continuous blackbody spectrum, but here on Earth many absorption

Focus Question 3-12

Diffraction gratings can either reflect or transmit light to create spectra. Are the peacock feathers in Figure 3-47b reflection or transmission diffraction gratings?

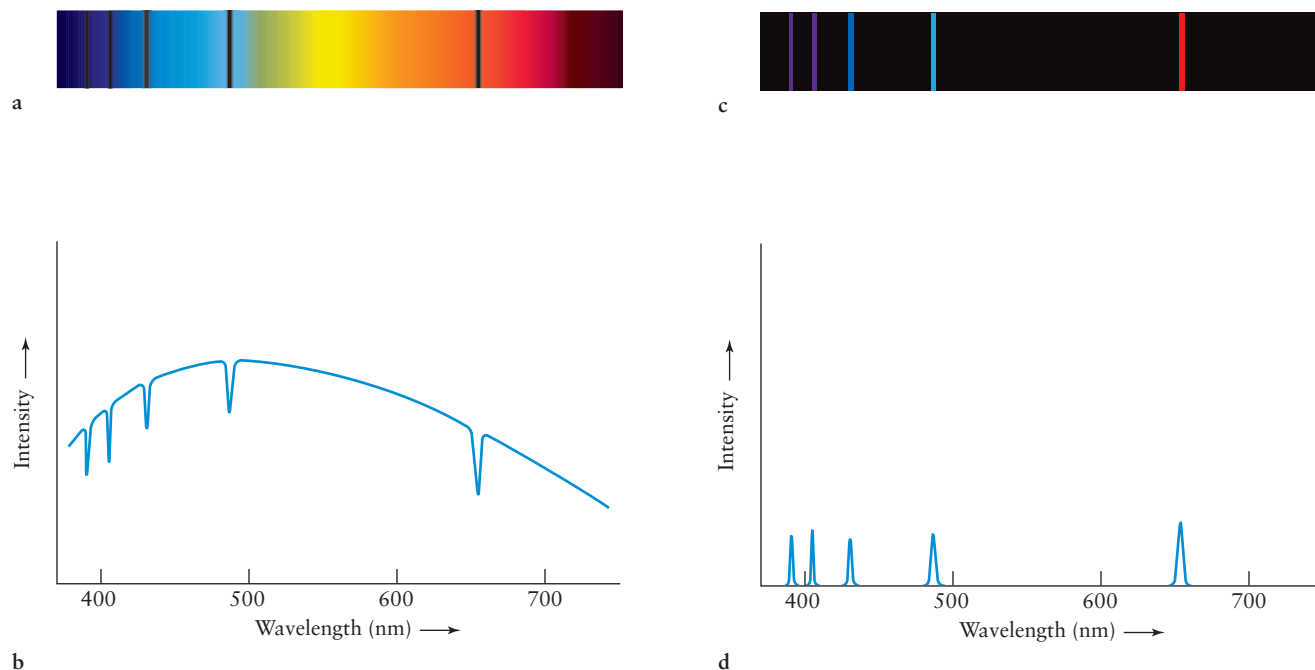


FIGURE 3-48 Spectrum of Hydrogen Gas (a) When a CCD is placed at the focus of a spectrograph, the spectrum is recorded. In this case, the spectrum is from a hot object whose light shines through hydrogen gas. (b) This spectrum is converted by computer into a graph of intensity versus wavelength. Note

that the absorption lines appear as dips in the intensity-versus-wavelength curve. Conversely, when a gas emits only a few wavelengths, (c) its emission spectrum appears as a series of bright lines, which is converted into peaks (d) on a graph of intensity versus wavelength.

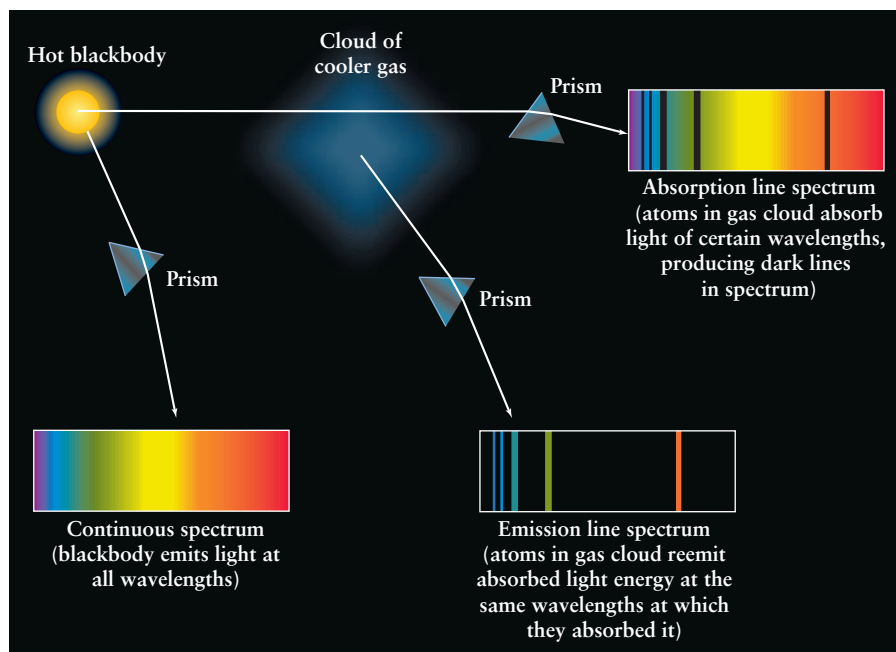


FIGURE 3-49 Continuous, Absorption Line, and Emission Line Spectra This schematic diagram summarizes how different types of spectra are produced. The prisms are added for conceptual clarity, but in real telescopes diffraction gratings are used to separate the colors. A hot, glowing object emits a

continuous spectrum. If this source of light is viewed through a cool gas, dark absorption lines appear in the resulting spectrum. When the same gas is viewed against a cold, dark background, its spectrum consists of just bright emission lines.

lines are seen in it (see Figure 3-44). Kirchhoff's third law explains why. There must be a cooler gas between the surface of the Sun and Earth. In fact, as noted earlier, there are two such gases: the Sun's lower atmosphere and Earth's entire atmosphere. Using the catalogs of spectra of different elements, the absorption lines in the Sun's spectrum tell us the chemical composition of the Sun's lower atmosphere and that of Earth's atmosphere.

If you have younger siblings or are around children a lot, you are likely to hear them go through a stage of development in which they frequently ask, "Why?" Scientists retain that curiosity, and the appearance of spectra is just one of a myriad of things about which they ask "why?" Fortunately, we can answer this question now, and the insights that it provides will be useful in comprehending concepts in virtually every chapter of this book. Our first step in understanding how spectra arise is to explore the building blocks of matter: atoms.

ATOMS AND SPECTRA

An atom is the smallest particle of a chemical element that still has the properties of that element. At the time of Kirchhoff's discoveries, scientists knew that all matter is composed of atoms, but they did not know how atoms were structured. Furthermore, scientists saw that atoms of a gas somehow extract light of specific wavelengths from continuum spectra that pass through the gas, leaving dark absorption lines, and they perceived

that the atoms then radiate light of precisely the same wavelengths—the bright emission lines (see Figure 3-46). But traditional theories of electromagnetism could not explain these phenomena. The answer came early in the twentieth century with the development of nuclear physics and quantum mechanics.

3-16 An atom consists of a small, dense nucleus surrounded by electrons

The internal structure of atoms first came into focus in 1908 when New Zealand native and 1908 Nobel laureate Ernest Rutherford and his colleagues at the University of Manchester in England were investigating radioactivity. They were expanding on work done earlier by Polish physicist (and *double* Nobel laureate, 1903 and 1911) Marie Curie, among others. Over time, a **radioactive** element naturally and spontaneously transforms into another element by emitting particles. Certain radioactive elements, such as uranium and radium, were known to emit particles moving with considerable speed. It seemed plausible that a beam of these high-speed particles would penetrate a thin sheet of gold. Rutherford and his associates found that almost all of the particles did pass through the gold sheet with little or no deflection. To their surprise, however, an occasional particle bounced back (**Figure 3-50**). It must have struck something very dense indeed.

Within a few decades, the structure of atoms became evident. That dense "something" is now called the nucleus

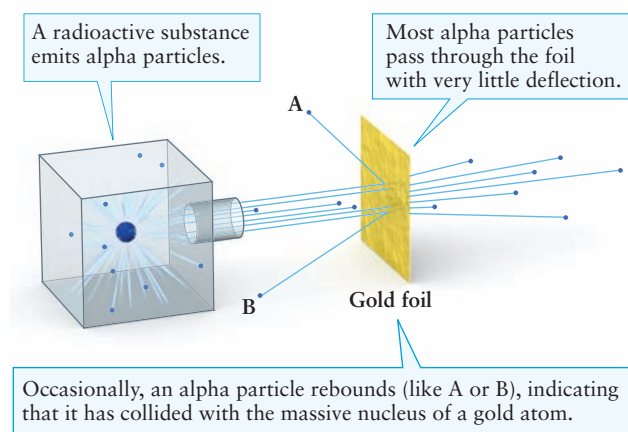


FIGURE 3-50 Rutherford Scattering Experiment Most helium nuclei (also called alpha particles) entering a thin foil scatter slightly as they pass through the medium, but some scatter backward, indicating that they have encountered very dense, compact objects. Such experiments were the first evidence that most matter is concentrated in what are now called atomic nuclei.

of the atom, and it consists of particles called protons and neutrons. Surrounding the nucleus, one or more electrons normally orbit. Newton's second law (see Section 2-7) explains that electrons move much more than the nuclei they orbit because the electrons have only 1/2000 the mass of a proton or neutron.

Unlike the planets, whose gravitational interaction keeps them orbiting the Sun, electrons orbit nuclei because electrons and protons have a property called *electric charge*. All protons have the exact same positive charge, while all electrons have a negative charge equal in strength to the proton's charge. The terms *positive* and *negative* are arbitrary and just indicate that the charges are opposite to each other. Particles with opposite charges attract each other. Therefore, protons attract electrons, and it is this attraction that keeps electrons in orbit around nuclei. The interaction between charged particles is a result of the **electromagnetic force**, the second of four fundamental forces in nature. (Gravitation is the first fundamental force we have discussed.)

Particles with the same type of charge, such as a pair of protons or a pair of electrons, repel each other. For atoms that have more than one proton in their nuclei (and many do), the protons are pushing away from each other. For nuclei to have more than one proton, there must be an attractive force stronger than the repulsion of protons to keep them glued together. The electrically neutral neutrons in the nucleus help provide that attractive force, called the **strong nuclear force**. This is the third of the four fundamental forces.

Many scientists find it fascinating that all of the interactions between matter and energy in nature occur as a result of just four forces, the above three plus the **weak nuclear force**. The weak nuclear force is involved

in some radioactive decays, such as when a neutron transforms into a proton. In Chapter 14, we will further explore the weak nuclear force in our study of the evolution of the whole universe. The properties of all four fundamental forces are summarized in **Table 3-2**.

The number of protons in an atom's nucleus, called the **atomic number**, determines the element of that atom. All hydrogen nuclei have 1 proton, all helium nuclei have 2 protons, all carbon nuclei have 6 protons, and so forth. There are 92 different types of elements that form naturally. Uranium is the most massive, with 92 protons (see Appendix N: Periodic Table of the Elements). As we will explore in Chapter 12, the fact that neutrons can transform into protons by emitting an electron and that protons can transform into neutrons by absorbing an electron is vital for the formation of elements with more than 26 protons.

In contrast to each atom of a given element always having the same number of protons in its nucleus, the nuclei of most elements can have different numbers of neutrons. For example, hydrogen, with 1 proton, can have 0, 1, or 2 neutrons; oxygen, with 8 protons, can have 8, 9, or 10 neutrons. Each different combination of protons and neutrons is called an **isotope**. There are three isotopes of hydrogen and three isotopes of oxygen. Hydrogen with no neutrons is the most common hydrogen isotope, while oxygen with 8 neutrons is by far the most abundant isotope of oxygen.

Some isotopes are stable, meaning that the numbers of protons and neutrons in their nuclei do not change. However, many elements have isotopes that are unstable and come apart spontaneously. These isotopes are radioactive, as discussed earlier in this section. For example, carbon with 6 neutrons, ^{12}C ("carbon twelve"), is stable, while carbon with 8 neutrons, ^{14}C ("carbon fourteen"), is unstable. ^{14}C decays into nitrogen with 7 neutrons, ^{14}N . Using radioactive age-dating techniques (see

TABLE 3-2 THE FOUR FUNDAMENTAL FORCES OF NATURE

Name	Strength (compared to the strong force)	Range of effect (from each object)
Strong force	1	Inside atomic nuclei
Electromagnetic force	1/137	Throughout the universe
Weak force	10^{-5}	Inside atomic nuclei
Gravitational force	6×10^{-39}	Throughout the universe

Appendix Q: Radioactivity and the Ages of Objects) on the oldest rocks brought back from the Moon, astronomers are able to determine roughly when the Moon formed (4.5 billion years ago). Applying the same technique to space debris found on Earth (meteorites), we can determine that the solar system, consisting of the Sun and everything orbiting it, formed some 4.57 billion years ago.

The number of electrons that orbit an atom is normally equal to the number of protons in its nucleus, thus making the atom electrically neutral. Astronomers denote neutral atoms by writing the atomic symbol followed by the Roman numeral I. For example, neutral hydrogen is written as H I and neutral iron is Fe I.

When an atom contains a different number of electrons than protons, the atom is called an *ion*. The process of creating an ion is called ionization. Ions are denoted by the atomic symbol followed by a Roman numeral that is one greater than the number of missing electrons. Positively ionized hydrogen (missing its one electron) is denoted H II, while positively ionized iron with seven electrons missing is denoted Fe VIII. Negative ions also exist, in which nuclei have more electrons orbiting than they have protons.

Atoms can share electrons and, by doing so, become bound together. Such groups of atoms are called molecules. Molecules are the essential building blocks of all complex structures, including life.

3-17 Spectra occur because electrons absorb and emit photons with only certain wavelengths

Because protons and neutrons have masses about 2000 times greater than the mass of an electron, over 99.95% of the mass of any atom is concentrated in its nucleus. The electron orbits are far from the nucleus, typically 10,000 times farther away than the radius of the nucleus. This immense distance is why you may have heard the statement that matter is mostly empty space.

Be careful not to imagine electrons as miniature planets orbiting the nucleus as a “miniature Sun.” Protons, neutrons, and electrons are not tiny solid bodies. Rather, like photons, they all have both wave and particle properties. The science that accurately describes their complex behavior is called **quantum mechanics**.

Quantum mechanics explains that electrons in atoms can exist in only certain *allowed orbits* around their nuclei, except when they are making a **transition** from one allowed orbit to another. These orbital conditions are completely unlike planets, which can exist at any distance (in any orbit) around the Sun. Each allowed electron orbit has a well-defined energy associated with it, and every different type of atom and molecule has a unique set of allowed orbits. These orbits and the transitions between them are the key to understanding the spectra that we have been discussing.

Consider an atom of the simplest hydrogen isotope, which contains just a single proton in its nucleus orbited by one electron. (This discussion generalizes directly to all of the other atoms and isotopes, made more complex only because they have more than one electron in orbit.)

Figure 3-51 shows a hydrogen atom’s lowest energy levels.

Normally, the hydrogen atom’s electron is in the lowest-energy allowed orbit or energy level, commonly called the ground state; this is labeled $n = 1$ in Figure 3-51. Each allowed orbit with successively higher energy is labeled $n = 2$, $n = 3$, $n = 4$, and so on. When an electron is in an orbit with more energy than the lowest-energy state available to it, it is said to be in an excited state. Electrons change orbits by absorbing or emitting photons. When a photon is absorbed, it vanishes and its energy becomes energy of the electron. However, electrons in orbit cannot absorb just any photon that they encounter. Electrons can absorb only those photons with energies exactly enough to boost them up to a higher-energy allowed orbit or with enough energy to completely remove them from orbit, an event called *photoionization*. Put another way, the nonionizing photons that are absorbed by electrons in atoms are those photons with energies equal to the difference between the energies of two allowed orbits. All nonionizing photons that do not satisfy this condition pass straight through the atom. If the electron starts in its ground state, then it must get exactly the energy necessary to move it to an excited state. If it is already in an excited

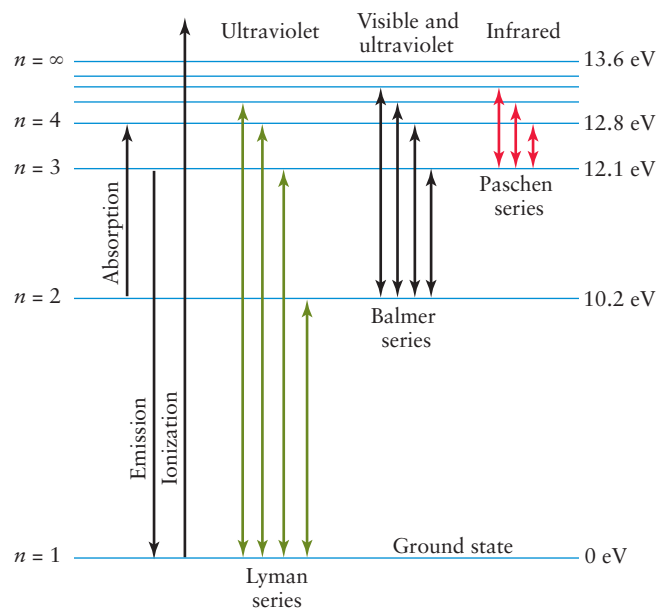


FIGURE 3-51 Energy Level Diagram of Hydrogen The activity of a hydrogen atom’s electron is conveniently displayed in a diagram showing some of the energy levels, labeled n , at which the electron can exist. Various electron jumps, or transitions, are also shown, including those that produce the most prominent lines in the hydrogen spectrum.

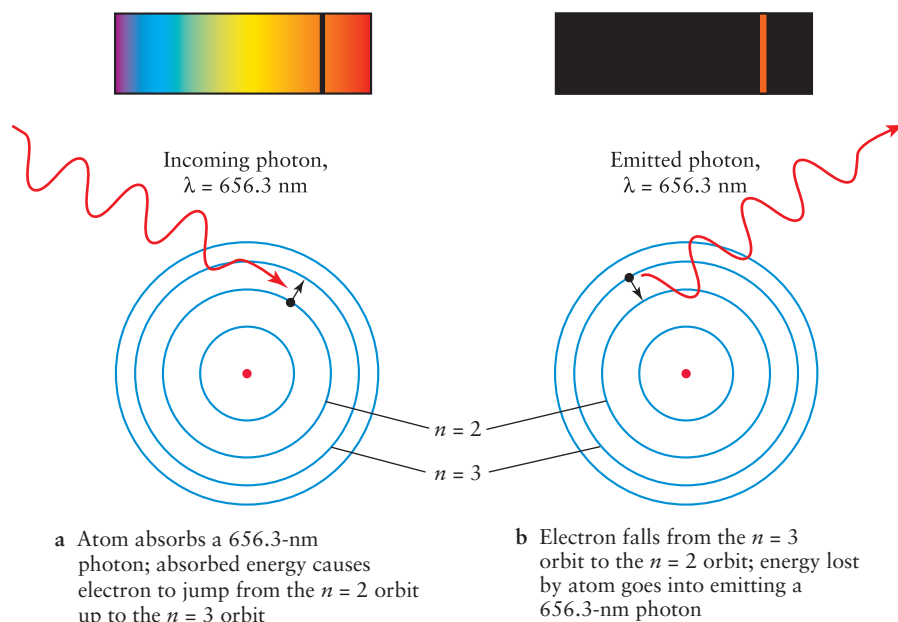


FIGURE 3-52 The Absorption and Emission of an H_α Photon This schematic diagram of hydrogen's four lowest allowed orbits shows what happens when a hydrogen atom absorbs or emits an H_α photon, which is red and has a wavelength of 656.3 nm. The spectral lines created by such events are also

shown. **(a)** A photon is absorbed by the electron, causing the electron to transition from orbit $n = 2$ up to orbit $n = 3$. **(b)** A photon is emitted as the electron makes a transition from orbit $n = 3$ down to orbit $n = 2$.

state, then by absorbing a photon, it must transition to a higher-energy excited state.

For example, referring to Figure 3-51 and **Figure 3-52**, transitions between the $n = 2$ and the $n = 3$ energy levels require the electron to absorb a photon with energy equal to $12.1 \text{ eV} - 10.2 \text{ eV} = 1.9 \text{ eV}$ (electron volt, a measure of energy). In this case, the photon absorbed by the electron has a wavelength of 656.3 nm (Figure 3-52a); it is a red photon (recall Figure 3-7).

Although some absorption occurs at optical wavelengths, it also happens in many other parts of the electromagnetic spectrum. Indeed, most of the transitions in hydrogen, shown in Figure 3-51 as the Lyman series (up from and down to $n = 1$) and the Paschen series (up from and down to $n = 3$), are nonvisible photons. Only a few of the Balmer series (up from and down to $n = 2$) transitions are visible, with the rest being ultraviolet.

We will refer to Balmer transitions when we study stars. They are named after the Swiss physicist Johann Balmer, who first calculated their wavelengths in 1885. The longest-wavelength Balmer line is called H_α (H alpha), the second H_β (H beta), the third H_γ (H gamma), and so forth, ending with the shortest-wavelength Balmer line, H_∞ (H infinity). (The first dozen lines of the series have Greek-letter subscripts; those remaining are identified by numerical subscripts.) Part of the spectrum of a star having Balmer absorption lines H_α through H_ζ (H zeta) is shown in **Figure 3-53**.

An electron in an excited state of hydrogen is unstable. It loses energy by bumping into other particles and suddenly having either too much or too little energy than is necessary to be in that state. When this happens, the electron is forced to descend to a lower-energy

Focus Question 3-13

Which Balmer line in Figure 3-51 is H_α ?

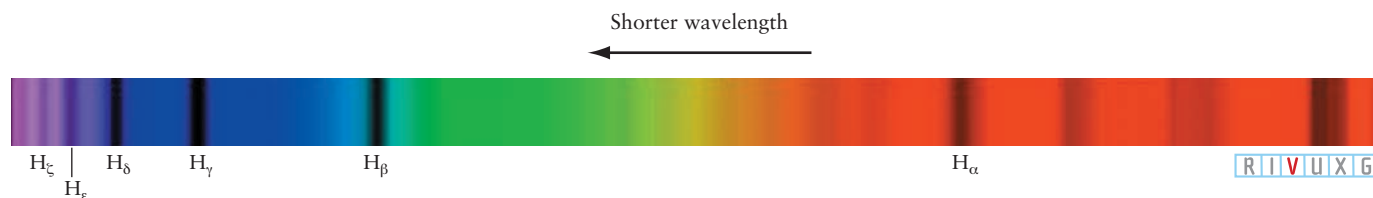


FIGURE 3-53 Balmer Lines in the Spectrum of a Star This portion of the spectrum of the star Vega shows six Balmer lines, from H_α at 656.3 nm through H_ζ at 388.9 nm. Absorption lines are therefore caused by photons being taken out of the stream of light by electrons, which thereby move into higher-energy

allowed orbits. An observer on the right in Figure 3-49 looking at the hot blackbody through the cloud of cooler gas sees dark absorption lines where photons have been absorbed by the cooler gas. (Chris and Dawn Schur, Schurastraphotography.com)

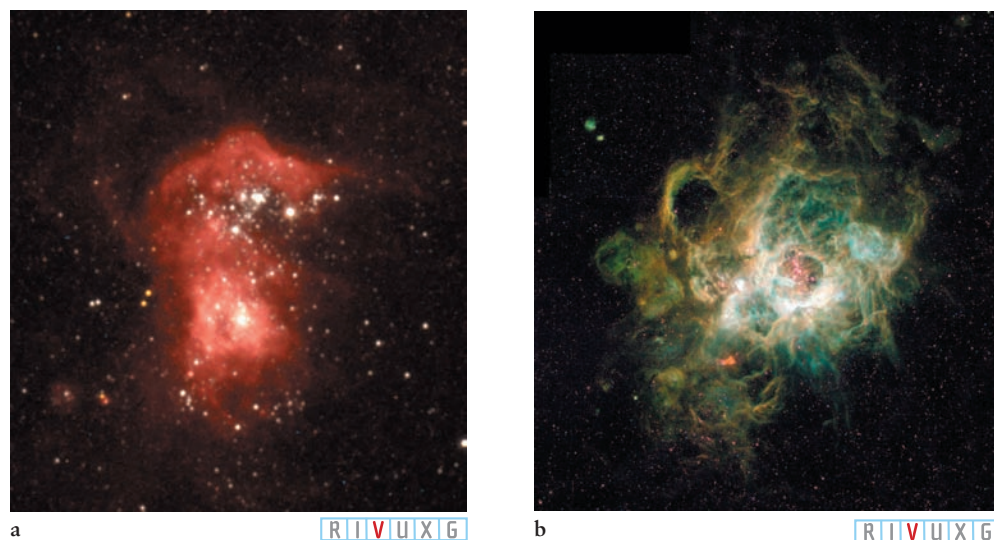


FIGURE 3-54 Emission Spectra from Interstellar Gas Clouds
(a) Stars in this interstellar gas cloud (NGC 2363) in the constellation Camelopardus (the Giraffe) emit blackbody spectra. Electrons in the cloud's hydrogen gas absorb and reemit the red light from these stars. NGC 2363 is located some 10 million ly away. **(b)** The nebula NGC 604, located 2.7 million ly away in the constellation Triangulum, is also an interstellar gas cloud.

The green glow is generated by doubly ionized oxygen atoms (O III; oxygen atoms missing two electrons) in the cloud that emit 501-nm photons. The Rosette Nebula is 3000 ly away. (a: Laurent Drissen, Jean-Rene Roy and Carmelle Robert [Department de Physique and Observatoire du Mont Megantic, Université Laval] and NASA and NSSDC; b: Hui Yang [University of Illinois] and NASA)

allowed state, either the ground state or some allowed state between where it starts and the ground state. In either case, to make this transition down, the electron must *emit* a photon with energy once again equal to the difference between the energies of the starting and final allowed orbits (see Figure 3-52b).

The emitted photons have the same set of wavelengths as the absorbed photons. Furthermore, the emitted photons are sent out (that is, scattered) in all directions, which creates the glowing emission line spectra of cool gases. The color of a gas depends on the atoms and molecules that it contains. Hydrogen-rich gas clouds glow red because of the H_α emission (Figure 3-53 and Figure 3-54a). Oxygen emits many green photons, so oxygen-rich gas clouds glow green (Figure 3-54b).

If an electron orbiting in a hydrogen atom encounters a photon with more than 13.6 eV, that photon is absorbed and knocks the electron completely out of orbit and away from the atom. As noted earlier in this section, this process is called photoionization. Each type of atom and molecule has a different photoionization energy, above which all electrons are kicked out of the atom or molecule. Photoionization occurs in stars and interstellar nebulae. Indeed, many spectral lines from stars and nebulae are those of ionized atoms that retain at least one electron, such as the doubly ionized oxygen atoms in Figure 3-54b.

3-18 Spectra provide information about motion of objects toward or away from us but not across the sky

The speed of an object toward or away from us is called its **radial velocity** (Figure 3-55) because the motion is along our line of sight or, put another way, along the “radius” drawn from Earth to the object. Of course, the star or other object may also have a velocity perpendicular or transverse to our line of sight, like a car passing a pedestrian. Both the radial velocity and the **transverse velocity** are measured in kilometers per hour.

Astronomers cannot measure the transverse velocity directly. Instead they measure the angle that the body moves among the more distant stars on the celestial sphere, which appear to move much less rapidly than do stars that are closer to us. This angle is called its **proper motion** and is often a matter of arcseconds per year or arcseconds per century. Proper motion does not affect the perceived wavelength of an object observed here on Earth and so proper motion cannot be determined by Doppler shift (recall section 3.2). The star with the greatest proper motion as seen from Earth is Barnard’s star (Figure 3-56). Planets moving among the relatively fixed stars on the celestial sphere are examples of bodies that have large proper motions.

Stellar proper motions are so small that they can be measured for only relatively nearby stars in our Galaxy.

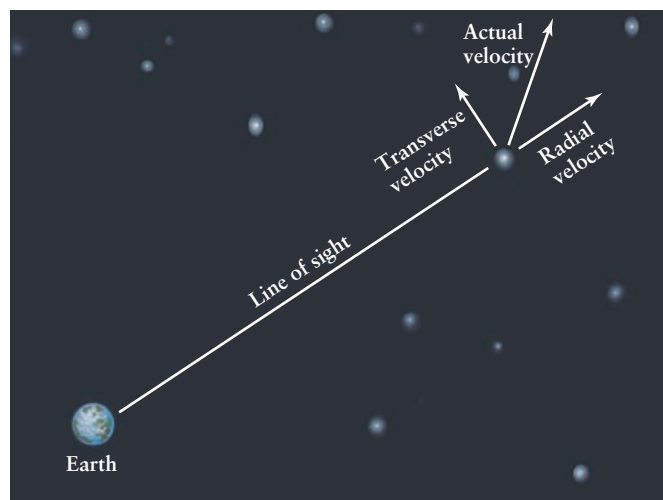
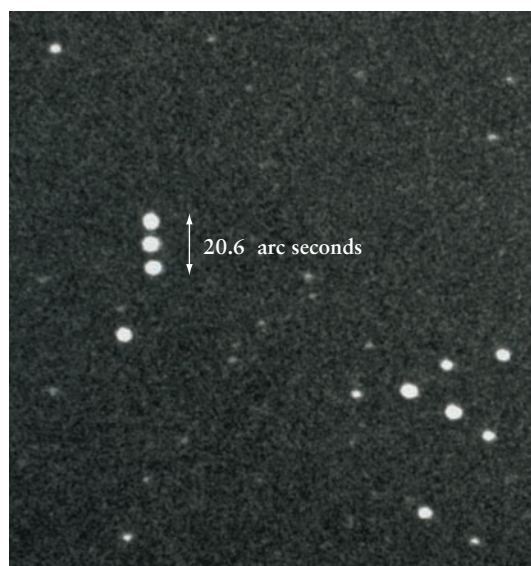


FIGURE 3-55 Radial and Transverse Velocities of a Star The actual velocity of a star (or other object) can be separated into radial and transverse motions. The speed of the star toward or away from Earth in kilometers per hour is its radial velocity. This motion creates a Doppler shift (see Section 3-2) of the entire star's spectrum. The speed in kilometers per hour perpendicular to the radial velocity is called its transverse velocity, which does not affect the star's spectrum. The *angular* motion of the star across the celestial sphere (that is, in the direction of its transverse velocity) is called proper motion.



RIVUXIG

FIGURE 3-56 Proper Motion of Barnard's Star Three images of Barnard's star, taken in 1975, 1977, and 1979, show the proper motion of the star during that time. In addition to having the largest known proper motion (10.3" per year), Barnard's star is one of the closest stars to Earth. (John Sanford/Science Source)

However, the radial velocity of virtually every object in space can be determined. Indeed, even the slow motions of rising and sinking gases on the Sun's surface can be measured by their spectral Doppler shifts. Doppler shift measurements of stars in double star systems give crucial data about the speeds of the stars orbiting each other. Doppler measurements of planets, the Sun, and other stars reveal that many of these bodies are rotating. Rotation is deduced when the Doppler shift shows that half of a surface is heading toward us, while the other half is simultaneously moving away from us. Furthermore, Doppler measurements reveal that all of the very distant galaxies are moving away from us, meaning that the universe is expanding. The spectra of distant galaxies enable us to determine the rate of that expansion. In later chapters, we will refer to the Doppler shift whenever we need to convert an observed wavelength shift into a speed toward or away from us.

With knowledge of spectroscopy, atomic and nuclear physics, and the Doppler shift, we can now list many of the vital properties of matter in space that are directly available to us from telescopic observations. These properties include the chemical compositions of stars, interstellar gas clouds, and other objects; the temperatures of these objects; the rotation of objects in space; the motion of objects toward or away from Earth; the presence of planets and dim companion stars; and the rate at which the universe is expanding. Knowing these properties, astrophysicists have developed and tested models that provide insights into the ages, masses, distances, internal activity, rotation rates, and companion objects of stars, among many other things.

Modern physics was born when Newton set out to understand the motions of the planets. Two and a half centuries later, scientists—including Maxwell, Planck, Einstein, Curie, Rutherford, and Doppler, among many others—discovered the basic properties of electromagnetic radiation and the structures of atoms. As we will see in the following chapters, the fruits of their labors have important implications for astronomy even today.

Insight Into Science

Remote Science Astronomical objects are so remote and their activities are often so complex that astrophysicists must use a tremendous amount of physics to interpret observations. For example, a single spectrum can contain information about stars, extrasolar planets, interstellar gas, Earth's motion and atmosphere, and the performance of the observing telescope and the equipment attached to it. All of these factors must be understood theoretically and accounted for.

SUMMARY OF KEY IDEAS

The Nature of Light

- Photons, packets of vibrating electric and magnetic fields, all carry energy through space at the same speed, the speed of light (300,000 km/s in a vacuum, slower in any medium).
- From longest to shortest wavelengths, radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays are all the forms of electromagnetic radiation.
- Photons behave sometimes as particles, sometimes as waves.
- Visible light occupies only a small portion of the electromagnetic spectrum.
- The wavelength of a visible-light photon is associated with its color. Wavelengths of visible light range from about 400 nm for violet light to 700 nm for red light.

Optics and Telescopes

- A telescope's most important function is to gather as much light as possible. When possible, it also resolves (reveals details) and magnifies an object.
- Reflecting telescopes, or reflectors, produce images by reflecting light rays from concave mirrors to a focal point or focal plane.
- Refracting telescopes, or refractors, produce images by bending light rays as they pass through glass lenses. Glass impurity, opacity to certain wavelengths, and structural difficulties make it inadvisable to build extremely large refractors. Reflectors are not subject to the problems that limit the usefulness of refractors.
- Earth-based telescopes are being built with active optics and adaptive optics. These advanced technologies yield resolving power comparable to the Hubble Space Telescope.

Capturing Nonvisible Light: Nonoptical Astronomy

- Radio telescopes have large, reflecting antennas (dishes) that are used to focus radio waves.
- Very sharp radio images are produced with arrays of radio telescopes linked together in a technique called interferometry.
- Earth's atmosphere is fairly transparent to most visible light and radio waves, along with some infrared and ultraviolet radiation arriving from space, but it absorbs much of the electromagnetic radiation at other wavelengths.
- For observations at other wavelengths, astronomers mostly depend upon space telescopes. Such satellite-based observatories are giving us a wealth of new information about the universe and permitting coordinated observation of the sky at many wavelengths.

- Charge-coupled devices (CCDs) record images on many telescopes used between infrared and X-ray wavelengths.
- By studying the wavelengths of electromagnetic radiation emitted and absorbed by an astronomical object, astronomers can learn about the object's temperature, chemical composition, rotation rate, companion objects, and movement through space.

Blackbody Radiation

- A blackbody is a hypothetical object that completely absorbs all the electromagnetic radiation that strikes it. The relative intensities of radiation at different wavelengths that it then emits depend only on its temperature. Stars closely approximate blackbodies.
- Wien's law states that the peak wavelength of radiation emitted by a blackbody is inversely proportional to its temperature—the higher its temperature, the shorter the peak wavelength. The intensities of radiation emitted at various wavelengths by a blackbody at a given temperature are shown as a blackbody curve.
- The Stefan-Boltzmann law shows that a hotter blackbody emits more radiation at every wavelength than does a cooler blackbody.

Identifying the Elements by Analyzing Their Unique Spectra

- Spectroscopy—the study of electromagnetic spectra—provides important information about the chemical composition of remote astronomical objects.
- Kirchhoff's three laws of spectral analysis describe the conditions under which absorption lines, emission lines, and a continuous spectrum can be observed.
- Spectral lines serve as distinctive “fingerprints” that identify the chemical elements and compounds comprising a light source.

Atoms and Spectra

- An atom consists of a small, dense nucleus (composed of protons and neutrons) surrounded by orbiting electrons. Atoms of different elements have different numbers of protons, while different isotopes have different numbers of neutrons.
- Quantum mechanics describes the behavior of particles and shows that electrons can only be in certain allowed orbits around the nucleus.
- The nuclei of some atoms are stable, while others (radioactive ones) spontaneously split into pieces.
- The spectral lines of atoms of a particular element correspond to the various electron transitions between allowed orbits of that element. Each orbit has a different energy level. When an electron shifts from one energy level to another, a photon of the appropriate

energy (and, hence, a specific wavelength) is absorbed or emitted by the electron.

- All different elements, isotopes, and molecules have different sets of spectral lines.
- When an atom has more protons than electrons, or vice versa, it is said to be charged. An atom loses an electron when the electron absorbs a sufficiently energetic photon, which rips the electron out of orbit.
- The motion of an object toward or away from an observer causes the observer to see all of the colors from the object blueshifted or redshifted, respectively. This effect is generically called a Doppler shift.
- Motion across the sky, called proper motion, yields no Doppler shift.

WHAT DID YOU THINK?

1 *What is light?* Light—more properly “visible light”—is one form of electromagnetic radiation. All electromagnetic radiation (radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays) has both wave and particle properties.

2 *Which type of electromagnetic radiation is most dangerous to life?* Gamma rays have the highest energies of all photons, so they are the most dangerous to life. However, ultraviolet radiation from the Sun is the most dangerous electromagnetic radiation that we commonly encounter.

3 *What is the main purpose of a telescope?* A telescope is designed primarily to collect as much light as possible.

4 *Why do all research telescopes use mirrors, rather than lenses, to collect light?* Telescopes that use lenses have more problems, such as chromatic aberration, internal defects, complex shapes, and distortion from sagging, than do telescopes that use mirrors.

5 *Why do stars twinkle?* Rapid changes in the density of Earth’s atmosphere cause passing starlight to change direction, making stars appear to twinkle. Seen from space, stars do not twinkle.

6 *Which is hotter, a “red-hot” or a “blue-hot” object?* Of all objects that glow visibly from heat generated or energy stored inside them, those that glow red are the coolest.

7 *What color does the Sun emit most brightly?* The Sun emits all wavelengths of electromagnetic radiation. The colors it emits most intensely are in the blue-green part of the spectrum. Because the human eye is less sensitive to blue-green than to yellow, and because Earth’s atmosphere scatters blue-green wavelengths more readily than longer wavelengths, we normally see the Sun as yellow.

Review Questions

1. Describe reflection and refraction. How do these processes enable astronomers to build telescopes?
2. Give everyday examples of refraction and reflection.
3. Which side of the secondary mirror in Figure 3-9 is coated with aluminum? Justify your answer.
4. Explain some of the advantages of reflecting telescopes over refracting telescopes.
5. What are the three major functions of a telescope?
6. What is meant by the angular resolution of a telescope?
7. Why will many of the very large telescopes of the future make use of multiple mirrors?
8. What is meant by adaptive optics? What problem does adaptive optics overcome?
9. Compare an optical reflecting telescope to a radio telescope. What do they have in common? How are they different?
10. Why can radio astronomers observe at any time of the day or night, whereas optical astronomers are mostly limited to observing at night?
11. Why must astronomers use satellites and Earth-orbiting observatories to study the heavens at X-ray wavelengths?
12. What are NASA’s four Great Observatories, and in what parts of the electromagnetic spectrum do (or did) they observe?
13. Why did Rømer’s observations of the eclipses of Jupiter’s moons support the heliocentric, but not the geocentric, cosmology?
14. A blackbody glowing with which of the following colors is coolest?

a. yellow	d. violet
b. red	e. blue
c. orange	
15. Of the following photons, which has the highest energy?

a. infrared	d. X-ray
b. radio waves	e. ultraviolet
c. visible light	
16. The spectrum of which of the following objects will show a redshift?

a. an object moving just eastward on the celestial sphere

- b. an object moving just northward on the celestial sphere
- c. an object moving directly toward Earth
- d. an object moving directly away from Earth
- e. an object that is not moving relative to Earth

17. What is a blackbody? What does it mean to say that a star appears almost like a blackbody? If stars appear to be like blackbodies, why are they not black?

18. What is Wien's law? How could you use it to determine the temperature of a star's surface?

19. What is the Stefan-Boltzmann law? How do astronomers use it?

20. Using Wien's law and the Stefan-Boltzmann law, state the changes in color and intensity that are observed as the temperature of a hot, glowing object increases.

21. What is an element? List the names of five different elements, and briefly explain what makes them different from each other.

22. How are the three isotopes of hydrogen different from each other?

23. Why do different elements have different patterns of lines in their spectra?

24. What is the Doppler shift, and why is it important to astronomers?

25. Explain why the Doppler shift tells us only about the motion directly along the line of sight between a light source and an observer but not about motion across the celestial sphere.

Got It?

26. Why do stars twinkle?

27. Why do all research telescopes use primary mirrors rather than objective lenses?

28. For the purpose of observing very faint objects, which of the following features of a telescope is most important? Explain your answer.

- a. its maximum magnification
- b. its ability to resolve colors
- c. the size of its objective lens or primary mirror
- d. the type of mount it has
- e. its weight

29. Of the following types of electromagnetic radiation, which is most dangerous to life?

- a. radio waves
- b. X-rays
- c. ultraviolet radiation
- d. infrared radiation
- e. visible light

30. What color does the Sun emit most intensely?

31. A star of which of the following colors is hottest?

- a. blue
- b. red
- c. orange
- d. yellow
- e. white

32. If a yellow star cools off, what color will it become next?

Advanced Questions

33. Advertisements for home telescopes frequently give a magnification for the instrument. Is this a good criterion for evaluating such telescopes? Explain your answer.

34. Show by means of a diagram why the image formed by a simple refracting telescope is "upside down."

35. Why does no major observatory have a Newtonian reflector as its primary instrument, whereas Newtonian reflectors are popular among amateur astronomers?

36. Why will many of the very large telescopes of the future have ultrathin primary mirrors?

37. What color will an interstellar gas cloud composed of hydrogen glow, and why?

38. Explain how the spectrum of hydrogen is related to the structure of the hydrogen atom.

39. Do we see all the colors from each star? Why or why not?

Discussion Questions

40. Discuss the advantages and disadvantages of using a relatively small visible-light telescope in Earth's orbit (for example, the 2.4-m Hubble Space Telescope) versus a large visible-light telescope on a mountaintop (for example, the 8.3-m Subaru Telescope on Mauna Kea, Hawaii).

41. If you were in charge of selecting a site for a new observatory, what factors would you consider?

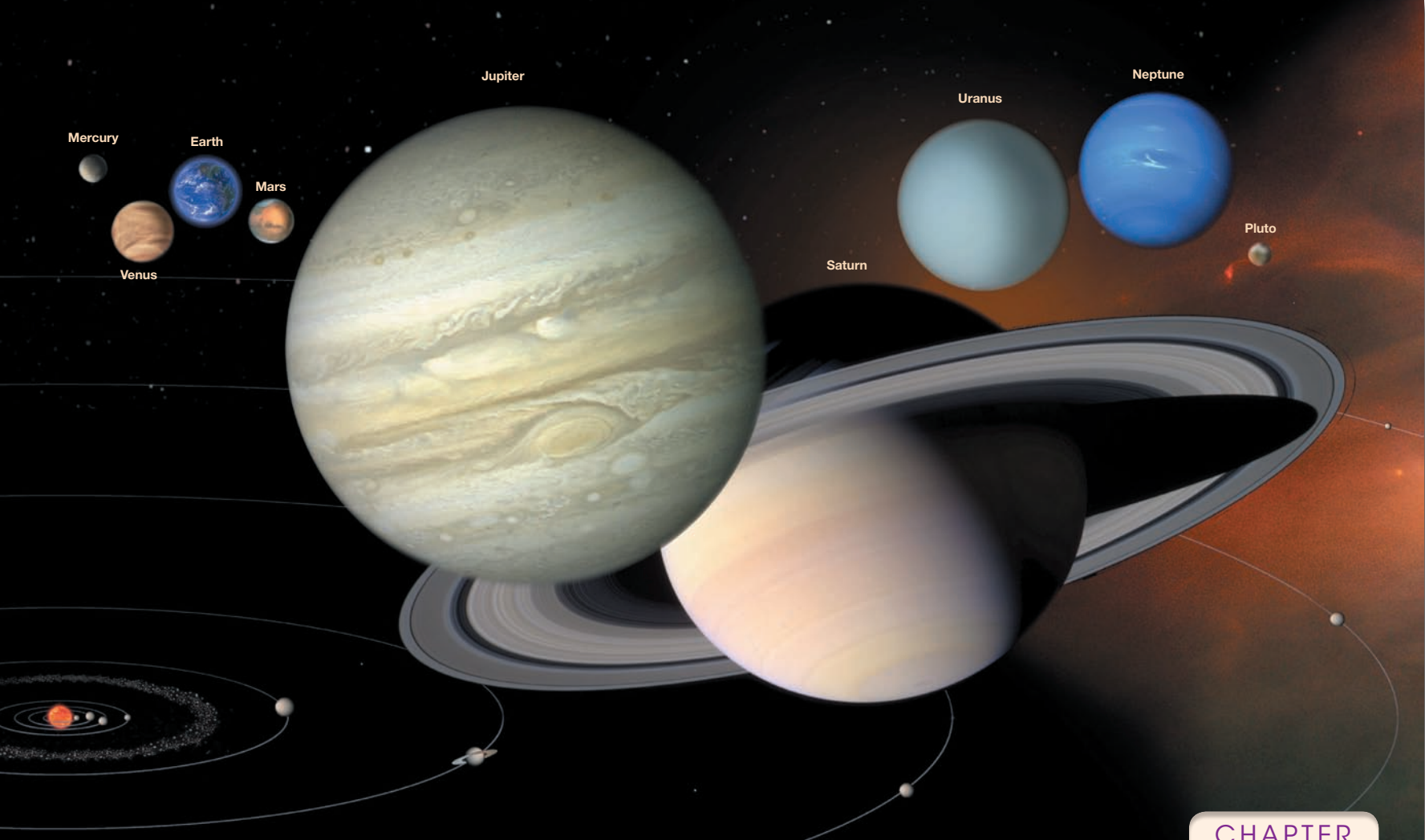
42. Consider two identical Cassegrain telescope mirrors. One is set up as a prime focus telescope, whereas the other is used in a Cassegrain telescope.

- a. Sketch both telescopes.
- b. What are the differences between the two that make each useful in different observing situations?

43. Compare the technique of identifying chemicals by their spectral line patterns with that of identifying people by their fingerprints.

44. Suppose you look up at the night sky and observe some of the brightest stars with your naked eye or binoculars. Is there any way of telling which stars are hotter and which are cooler? Explain your answer.

45. How could we exclude Earth's atmosphere as the source of the iron absorption lines in Figure 3-46?



R I V I U X I G A montage of the planets (plus Pluto, a dwarf planet) in our solar system presented in correct relative sizes. The orbits in the background are also drawn to scale. (NASA)

CHAPTER 4

Formation of the Solar System

WHAT DO YOU THINK?

- 1 How many stars are there in the solar system?
- 2 Were the Sun and planets among the first generation of objects created in the universe?
- 3 How long has Earth existed, and how do we know this?

Answers to these questions appear in the paragraphs beside the corresponding circled numbers in the margins and at the end of the chapter.

In this chapter you will discover

- how the solar system formed
- why the environment of the early solar system was much more violent than it is today
- how astronomers define the various types of objects in the solar system
- the relationships between planets, dwarf planets, small solar system bodies, and other classifications of objects in the solar system
- how and why the planets are grouped as they are
- how the moons formed throughout the solar system
- the composition of the debris scattered throughout the solar system

THE SOLAR SYSTEM CONTAINS HEAVY ELEMENTS, FORMED FROM AN EARLIER GENERATION OF STARS

1 The **solar system** is made up of one star, our Sun, and all the bodies that orbit it—planets, moons, and various kinds of debris (called asteroids, comets, and meteoroids). How did the solar system form? How were its varied building blocks of rock, metal, ice, and gas created? How has the solar system changed since its formation, and what does its history tell us about the planets as they are today? Within the past few decades, telescopes and space probes, along with the theories of modern science and computer simulations of these theories, have finally provided some of the answers to these age-old questions.

4-1 Stars transform matter from lighter elements into heavier ones

Our study of the formation of the solar system begins with an inconsistency, namely, that Earth, the Moon, and many other objects that orbit the Sun are composed primarily of heavy elements including, among many others, oxygen, silicon, aluminum, iron, carbon, and calcium. (Note that we capitalize the word *moon* only when we are referring to Earth's Moon.) Furthermore, most of these bodies contain extremely little hydrogen and helium. However, observations of the spectra of the Sun, other stars, and interstellar clouds reveal that hydrogen and helium are by far the most abundant elements in the universe. These two elements account for about 99.9% of all observed atoms (or, equivalently, 98% of observed mass). All of the other elements combined account for about 0.1% of the observed atoms (and, hence, 2% of the observed mass of the universe). How is it that Earth, the Moon, Mars, Venus, Mercury,

and many smaller bodies in the solar system typically contain less than 0.15% hydrogen and helium? Somehow, these familiar astronomical objects formed from matter that had been enriched with the heavier elements and depleted in hydrogen and helium.

There is a good reason for the overwhelming abundance of hydrogen and helium throughout the universe. Astronomers believe that the universe formed about 13.8 billion years ago in a violent event called the *Big Bang*, which created not only all the matter and energy that exists but also space and time. (We explore these issues in Chapter 14.) Only the lightest elements—hydrogen, helium, and a tiny amount of lithium—were created as a direct result of the Big Bang. These first elements came into existence as gases. The first stars, composed only of these three elements, condensed out of this primeval matter within a few hundred million years after the Big Bang occurred.

2 The difference in chemical composition between the gas and stars in the early universe and the solar system today shows that the solar system did not form as a *direct* result of the Big Bang. Rather, as we shall see shortly, the solar system formed some 9 billion years after the universe came into existence. Let us briefly explore the chemical transformations that led to our being here, which began deep inside those first stars.

Once stars form, gravity compresses the matter in their central *cores* so much that the hydrogen there is transformed into helium in a process called *thermonuclear fusion* (*fusion*, for short). Hydrogen fusion has the interesting property that some of the hydrogen's mass is converted into electromagnetic energy (photons), a portion of which eventually leaks out through the star's surface, enabling it to shine. (That is why the Sun shines.) The core's helium subsequently fuses to create carbon and oxygen, also converting more mass into energy. If the star has enough mass, the pressure in its core will force these elements to transform into even heavier elements, including neon, aluminum, silicon, calcium, and iron, while other elements are fused in shells surrounding the core. We explore details of

Focus Question 4-1

Which elements on Earth may have been unchanged since the universe began?

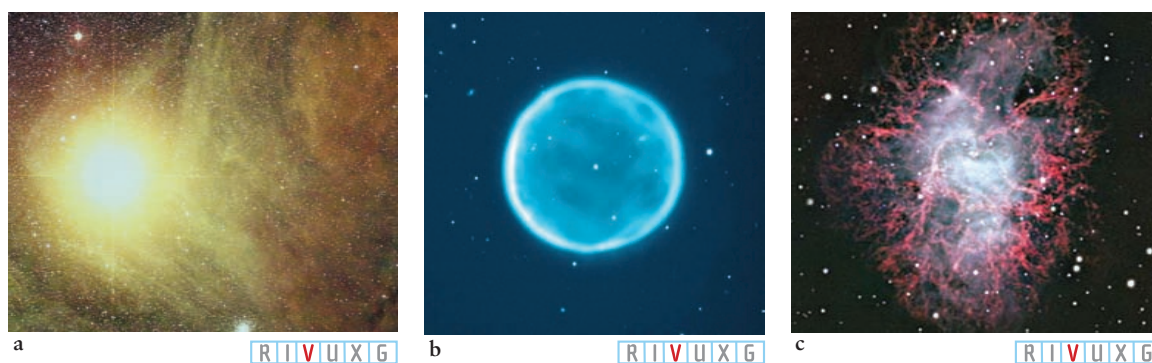


FIGURE 4-1 How Stars Lose Mass (a) Antares, the brightest star in Scorpius, is nearing the end of its existence. Strong winds from its surface are expelling large quantities of gas and dust, creating this nebula reminiscent of an Impressionist painting. The scattering of starlight off this material makes it appear especially bright, even at a distance of 604 ly. (b) The planetary nebula Abell 39 is 7000 ly from Earth. With a relatively gentle emission of matter, the central star

shed its outer layers of gas and dust in an expanding spherical shell now about 6 ly across. (c) A supernova is the most powerful known mechanism for a star to shed mass. The Crab Nebula, even though it is about 6000 ly from Earth, was visible during the day for three weeks during 1054. (a: Australian Astronomical Observatory/David Malin Images; b: George Jacoby [WIYN Obs.] et al., WIYN, AURA, NOAO, NSF; c: Malin/Pasachoff/Caltech)

the creation of these heavier elements in Chapters 11 and 12.

Stars also shed matter into space (Figure 4-1). Some of the heavy elements that formed inside them, specifically those elements created in shells of fusion occurring outside their cores, are eventually ejected, along with most of the hydrogen and helium remaining in their outer layers. The outer layers of stars are expelled at different rates, ranging from continuous outflows called *stellar winds* (Figure 4-1a), to the more energetic expulsion of gases called *planetary nebulae* (Figure 4-1b), to spectacular detonations called *supernovae* (Figure 4-1c). Planetary nebulae were so named not because they have any direct connection with real planets, but rather because they looked like planets in early, low-resolution telescopes. Supernovae are such powerful events that fusion occurs during these explosions, creating a variety of heavy elements, such as nickel, copper, zinc, silver, and gold, which are also expelled from the star. Planetary nebulae and supernovae leave only tiny, dim, but in many cases very hot stellar cores, remnants of once mighty stars.

Of the matter ejected from stars, some leaves as small particles of dust, similar to the soot formed in a fire or in diesel exhaust. This dust is typically a millionth of a meter (a micron) in size. Over time, enough gas and dust were emitted by enough stars to form interstellar clouds rich in hydrogen and helium that also contained small amounts of metals. (In a curious alternative use of a common word, astronomers define **metals** to be all of the elements in the universe other than hydrogen and helium.) Some of the atoms in these clouds also combine to form molecules of, for example, water, carbon dioxide, methane, and ammonia. We will turn next to these

clouds in order to understand the formation of the solar system and why dense planets like Earth are so rich in metals and poor in light elements.

4-2 Gravity, rotation, collisions, and heat shaped the young solar system

Observations of clouds, such as the Cone Nebula and the nearby Orion Molecular Cloud (Figure 4-2), reveal that most stars form from gas and dust inside such giant interstellar clouds. Between a few and a few thousand stars typically form at the same time in these clouds, creating groups called *open clusters* of stars. It is likely that the solar system formed in such a cluster. Stars in open clusters eventually drift apart, which is why even the closest stars to the solar system are light-years away from us today.

Such clouds do not *spontaneously* collapse to form stars because the particles in these clouds are moving fast enough (meaning they are warm enough, since motion is a measure of how hot gases are) to avoid being pulled together by their mutual gravitational attraction. Another way to say this is that the gases comprising these clouds have high enough internal pressures to prevent them from collapsing together and forming stars. Pressure is the force that the gas exerts on any area of either itself or anything else (think of the pressure that the gas in a balloon exerts on the surface of the balloon to keep it inflated).

With sufficiently low temperature and pressure and sufficiently high gas density, however, fragments of clouds collapse under the influence of their own gravitational attraction. Density is the amount of mass of any substance within a given volume. We say that

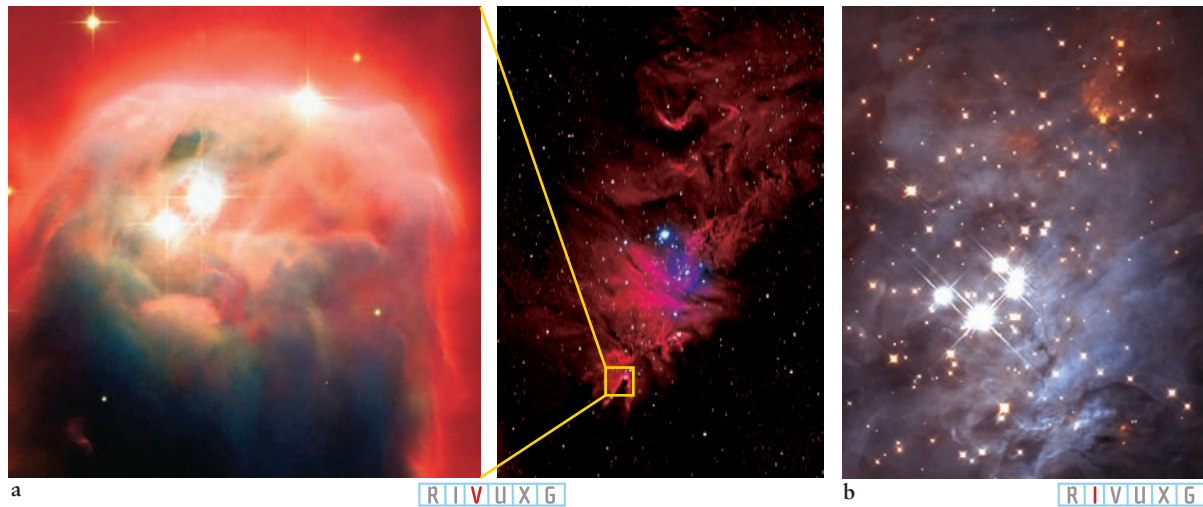


FIGURE 4-2 Dusty Regions of Star Formation (a) The three bright young stars shown in the inset of this image of the Cone Nebula in the constellation Monoceros are still surrounded by much of the gas and dust from which they formed. Astronomers hypothesize that the solar system formed from a similarly small fragment of a giant interstellar gas and dust cloud. (b) Newly formed stars in the Orion Nebula. Although visible

light from many of the stars is blocked by the nebula, its infrared emission travels through the gas and dust to us. (a: [left] Laura Baird/Flynn Haase/NOAO/AURA/NSF [right] ACS Science & Engineering Team/NASA; b: NASA/K. L. Luhman [Harvard-Smithsonian Center for Astrophysics] and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thomson [Steward Observatory, University of Arizona])

each of these regions has a **Jeans instability**, after the British physicist Sir James Jeans, who calculated the necessary conditions for such collapse in 1902. Pieces of interstellar clouds become “Jeans unstable,” allowing stars and their hosts of orbiting bodies to form, for at least three reasons. First, winds from nearby stars compress gas and dust in the cloud. Second, the explosive force of a nearby supernova compresses regions of gas and dust. Third, pairs of clouds collide and compress each other. Each collapsing fragment destined to become a star (or a pair of stars orbiting each other called a *binary star system*) and possibly planets is called a **dense core**.

Some chemical evidence in space debris that we have analyzed on Earth suggests that the dense core in which our solar system formed may have become Jeans unstable when its gas and dust were hit by the shock wave from a supernova. We call that dense core the **solar nebula** (also called the protosolar or presolar nebula).

Many of the concepts in what follows comprise the **Nice model** (pronounced like “niece” and named after the city in France where it was first developed) of the formation of the solar system. This model is based on computer simulations of the evolution of the outer regions of the solar nebula. *We are in an epoch of astronomical research in which insights into the process by which the solar nebula transformed into the Sun and its host of orbiting bodies are coming at an unprecedented rate. The details of this theory are tentative and are likely to evolve as simulations become more realistic and more solar system debris left over*

from the formation process is analyzed. Therefore, the following scenario is a prime example of how science is a work in progress rather than a presentation of complete understanding of nature.

The solar nebula initially had a diameter of at least 1000 AU (1000 times the average distance from Earth to the Sun) and a total mass about 2 to 3 times the solar system’s present mass (Figure 4-3). At first that nebula was a very cold collection of gas and dust—well below the freezing point of water. Although most of the solar nebula was hydrogen and helium, ice and ice-coated dust grains composed of heavy elements were scattered across this vast volume. Deep inside the nebula, gravitational attraction caused gas and dust to fall rapidly toward its center. (In terms of the physics presented in Appendix P: Energy and Momentum, the cloud’s gravitational potential energy was being converted into kinetic energy—energy of motion.) As a result, the density, pressure, and temperature at the center of the nebula began to increase, producing a concentration of matter called the **protosun**.

As the protosun continued to increase in mass and to contract, atoms within it collided with increasing speed and frequency. This activity created growing pressure inside the protosun, which caused its contraction to slow dramatically. Such collisions also created heat, causing the protosun’s temperature to soar. Therefore, the first heat and light emitted by our Sun came from colliding gas, not from thermonuclear fusion, as it does today. At the same time, there was also considerable activity in the outer regions of the disk.

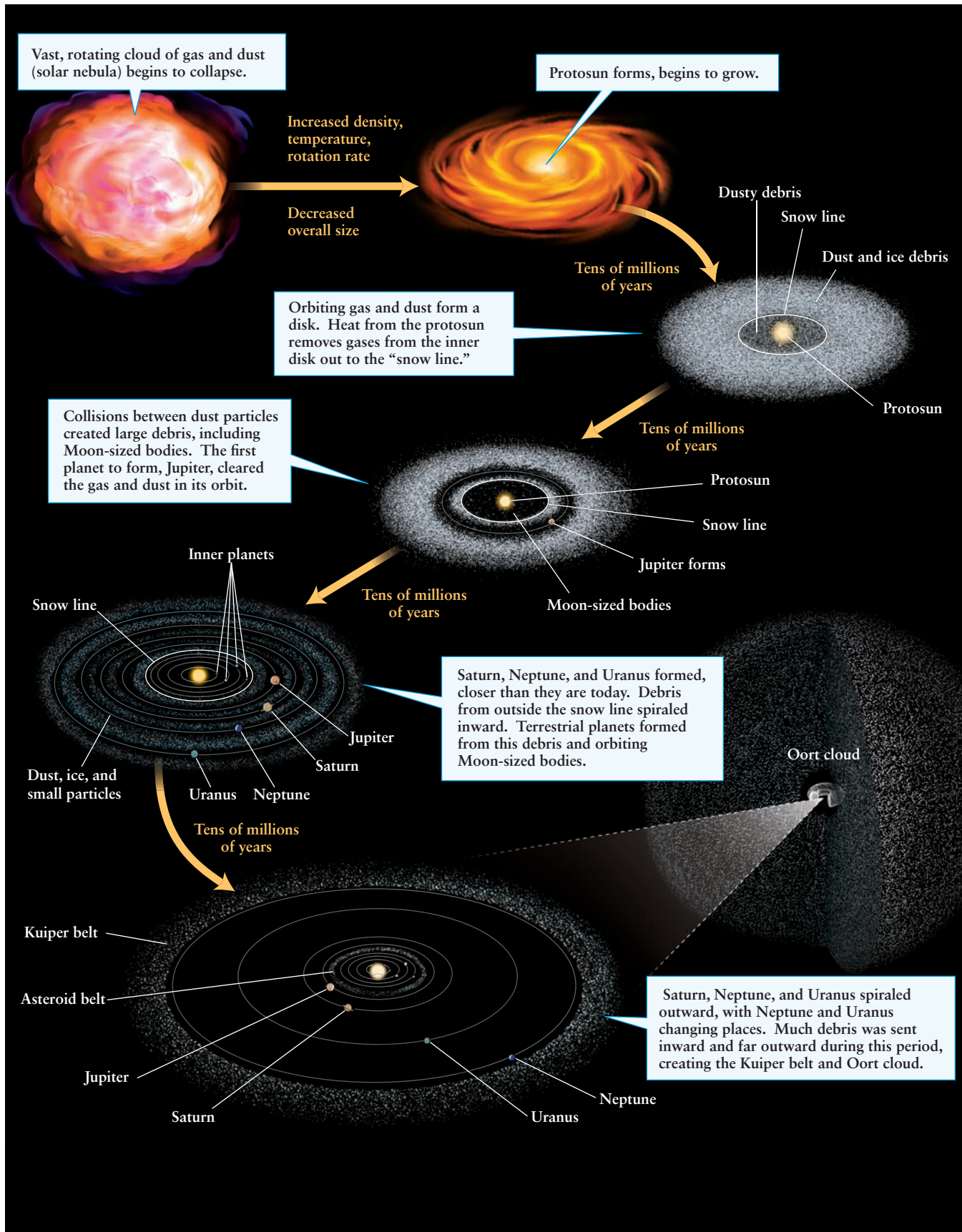


FIGURE 4-3 The Formation of the Solar System This sequence of drawings shows stages in the formation of the solar system.

Insight Into Science

Computer-Aided Analysis Many equations are so complex that they require computer analysis to enable us to understand their implications. For example, the study of the physics of the formation of the planets and the Sun began with observations, which then led to the Nice model, which uses ever more sophisticated computer simulations to reveal the early evolution of the solar system.

Rotation played a key role in the formation of the solar system. When it first developed, the solar nebula was a very slowly rotating ensemble of particles. This rotation occurred because the formation of interstellar clouds is turbulent, so these clouds are created with many slowly swirling regions, like smoke rising from a fire. As a result of this rotation, the solar nebula, created from one of these slowly swirling regions, had angular momentum (see Appendix P: Energy and Momentum). Because angular momentum is conserved (see Section 2-7), rotating matter spiraling inward in the young solar nebula revolved faster and faster, like the skater in Figure 2-12 when she pulls her arms in. This motion had the effect of flattening the nebula into a disk. You can see an analogous effect of rotation in the way the ice skater's dress in Figure 2-12 is forced into a plane as she spins. (Had the solar nebula been nonrotating, it would have collapsed en masse to create a star without any planets or other orbiting matter.) Even the debris in the inner solar nebula that formed the protosun was orbiting as it fell inward, like water spiraling before it goes

down a drain. Therefore, the protosun was spinning and the Sun that formed from it rotates. We will discuss the Sun's rotation further in Chapter 9.

The outer region of the young solar nebula was probably very ragged, but mathematical studies show that the combined effects of gravity, collisions, and rotation would transform even an irregular dense core into a rotating disk with a warm center and outer cold edge about 30 AU from the protosun (see Figure 4-3). Eventually, the gas and dust in the disk of the solar nebula were orbiting fast enough to stop spiraling inward.

Although we cannot see our solar system as it was before planets formed, astronomers have found what we think are similar disks of gas and dust surrounding other young stars, including those shown in Figure 4-4. Called **protoplanetary disks**, these systems are undergoing the same initial stage of evolution that we have described here for our solar system.

As occurs with all gases, the protosun's temperature increased as it became denser. As the protosun radiated more and more heat, the temperature around it began to increase. The protosun's increasing temperature vaporized all common icy substances (including carbon dioxide, water, methane, and ammonia) in the inner region of the disk. These gases then blew away, along with the abundant hydrogen and helium gas that also initially existed there. Within a hundred thousand years or so of the disk forming, no gases or ices orbited closer than roughly 3 AU, between what are now the orbits of Mars and Jupiter. Temperatures were low enough beyond this distance that gases, primarily hydrogen and helium, and ices remained in orbit around the protosun. The boundary beyond which these gases and ices of water, carbon dioxide, methane, and ammonia persisted is called the

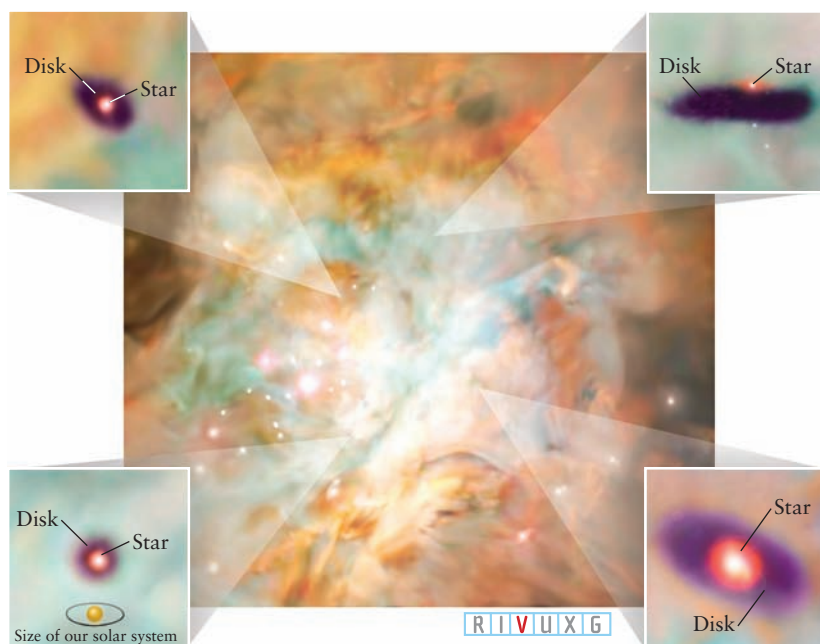


FIGURE 4-4 Young Circumstellar Disks of Matter The heart of the Orion Nebula as seen through the Hubble Space Telescope. The four insets are false-color images of protoplanetary disks within the nebula. A recently formed star is at the center of each disk. The disk in the upper right is seen nearly edge-on. Our solar system is drawn to scale in the lower left image. (C. R. O'Dell and S. K. Wong, Rice University/NASA)

snow line (or *frost line* or *ice line*). Within a million years of when the Jeans instability occurred, the solar system had differentiated into an inner region containing just metal-rich dust grains (that are not easily vaporized) and an outer region rich in hydrogen gas, helium gas, and ice-covered dust particles.

Neighboring dust grains beyond the snow line began colliding with each other. The early motion of this debris was very chaotic; many of these collisions caused dust grains to grow, while many other collisions broke apart larger pieces of this debris. When dust particles finally reached sizes of a few millimeters, they began plowing through the surrounding gas. As the dust particles pushed aside this gas, the gas took energy away from (that is, slowed down) the dust particles, much as you slow down when walking into a headwind. As a result of losing energy (slowing down), this dust spiraled inward toward the protosun. The collision and **accretion**—the coming together of smaller pieces of matter to form larger ones—of dust particles continued as they spiraled inward. Like static cling between clothes, accretion of dust-sized particles is due to the attraction of their charged particles—protons and electrons—while accretion of much larger pieces of debris is primarily due to their mutual gravitational attraction. Through accretion, the dust grew into pieces of debris meters across. Calculations show that a piece of debris a meter across could have spiraled inward to half its original distance from the protosun in just 1000 years! Collisions destroyed some of these bigger bodies, while some of them merged and grew even larger. Eventually, a few billion of these boulders reached dimensions of a kilometer or so, large enough for their mutual gravitational attractions to enhance the rate at which they collided. Such pieces of debris are called **planetesimals**.

The formation of planetesimals that had migrated inside the snow line was a turning point in the young solar system's evolution. They collided with countless dust particles as well as with each other. Sometimes colliding planetesimals destroyed each other, and returned to dust. Eventually, however, these collisions led to the growth of several hundred substantial bodies that persisted and eventually reached the size and mass of our Moon. At this size, the gravitational attraction of the planetesimals was strong enough for them to either sweep up the surrounding rubble or to fling it far away, thereby clearing out nearby space and limiting their own growth. These bodies led to the formation of the terrestrial planets, to which we will return shortly.

THE FORMATION OF THE PLANETS

4-3 The giant planets formed in sequence

Jupiter Many planetesimals that formed out beyond the snow line remained out there in substantially noncircular

orbits. As a result, they crossed paths and a few of them collided and merged with each other, creating larger bodies. One planetesimal located 1 or 2 AU beyond the snow line accreted enough of its companions to become an Earth-like body containing perhaps 10 Earth masses of rock and metal. This body had enough gravitational attraction to pull onto itself vast amounts of the hydrogen, helium, and ices that existed near its orbit. Within only a few thousand years of forming, it grew into the planet Jupiter. Like the heating of the protosun, impacts from infalling gases heated Jupiter up; in fact, the planet was heated so much that for a short time it outshone the protosun, if those two bodies had been observed then from equal distances. Jupiter, however, never became a star because its total mass is too low for it to compress and heat its core enough to enable hydrogen fusion to commence. Indeed, Jupiter would need to be 75 times more massive for it to sustain core fusion and therefore be classified as a star, as discussed in Section 4-1.

Plowing through the surrounding gas and dust, the growing Jupiter lost energy and was thereby forced to spiral inward. Calculations suggest that it may have reached as close as Mars's present orbit before it stopped its migrating toward the protosun. On the way in, Jupiter's gravitational attraction and the heat it emitted stirred up the surrounding gas enough so that some of the gas and dust it passed was sent inward toward the protosun and some was sent outward. In this way, Jupiter eventually "cleared its neighborhood," and its growth halted. By this point, Jupiter had accumulated about 300 Earth masses of hydrogen and helium, plus several Earth masses of water and other simple molecules. From the inside out, Jupiter was a rocky world, surrounded by a layer of water, which was surrounded in turn by a much thicker layer of predominantly hydrogen and helium. Jupiter rotates today because it formed from swirling debris in the solar nebula. Likewise, all the other planets in our solar system rotate.

Jupiter's inward migration had two long-term effects on our solar system. First, as it passed through what is now the asteroid belt, it dispersed most of the debris there, preventing a planet from ever forming in that region. Second, it had a similar, but not quite as effective, influence on stirring up the debris in the vicinity of what is now Mars's orbit. Astronomers believe that is why Mars is such a low-mass planet compared to Earth and Venus. As we will discuss in Chapter 5, many Jupiterlike planets orbiting other stars are even closer to their stars than Mercury is to the Sun. Jupiter is likely to have continued inward as it plowed through the surrounding gas and dust if it had not been stopped by the gravitational attraction of the next planet to form, Saturn.

Focus Question 4-2

What force ensured that once the solar nebula began collapsing, it continued to do so?

Saturn A few million years after Jupiter formed, collisions of planetesimals led to the formation of a second rocky world of about 10 Earth masses located a few astronomical units out beyond Jupiter. This second world underwent the same process of accreting hydrogen, helium, and ice, thereby creating Saturn. By the time Saturn formed, however, a lot of gas in its vicinity had already been stirred up and dissipated by Jupiter, leaving Saturn with less matter to accrete. Saturn has about 80 Earth masses of hydrogen and helium, along with a few Earth masses of water. Note that when it formed, Saturn was closer to the protosun than it is to the Sun today. Furthermore, the gravitational attraction of the young Saturn, combined with that of Jupiter, cleared their neighborhoods of the gas and dust that were causing them to spiral inward. As a result of their gravitational interaction with each other, they forced each other to spiral back out until Jupiter arrived at its present location, with Saturn farther out.

Focus Question 4-3

What two elements comprise most of Saturn's outer layer?

Neptune and Uranus The gravitational effects of Jupiter and Saturn forced many planetesimals to become concentrated out beyond Saturn's orbit, leading to the formation of two more rocky planets destined to become the cores of Neptune and Uranus. The debris that far from the protosun had been spiraling inward for several million years; thus, these two planets are believed to have formed substantially closer to Saturn than they are today. Furthermore, computer simulations suggest that at the time of their formation, Neptune was closer to the protosun than Uranus (the opposite of their positions now!).

While substantial ice remained for Neptune and Uranus to collect, they formed so late that Jupiter and Saturn had dissipated most of the hydrogen and helium

in their neighborhoods. Therefore, Neptune and Uranus became rich in water, along with carbon dioxide, methane, and ammonia—generically called “ices”—from their surroundings, but each accreted only a couple of Earth masses of hydrogen and helium. While Jupiter and Saturn are called *gas giants*, Neptune and Uranus are called *ice giants*.

4-4 The inner planets formed primarily from collisions

The details of how the inner four rocky planets—Mercury, Venus, Earth, and Mars—formed are less well known than for the giant planets just discussed. It appears likely that the inner planets formed as a result of collisions between the Moon-sized and larger planetesimals discussed at the end of Section 4-2. Within 100,000 years of the dust beginning to coalesce and to spiral inward, as described above, bodies with 0.1 Earth masses were able to form in it and collide with each other inside the snow line. It is unknown whether most of these planetesimals formed inside the snow line in relatively circular orbits and then collided to build planets in relatively circular orbits, or whether the four inner worlds formed from collisions of planetesimals that came from beyond the snow line in highly elliptical orbits that then somehow became circular. Whether the inner planets formed in orbits near their present average distances from the Sun or whether they migrated inward, like the planetesimals outside the snow line described above, is also unknown.

In the end, however, the collisions of hundreds of Moon-sized planetesimals led to the existence of four rocky inner planets: Mercury, Venus, Earth, and Mars (**Figure 4-5**). *We are literally made of stardust!* Because these planets were located inside the snow line, they did

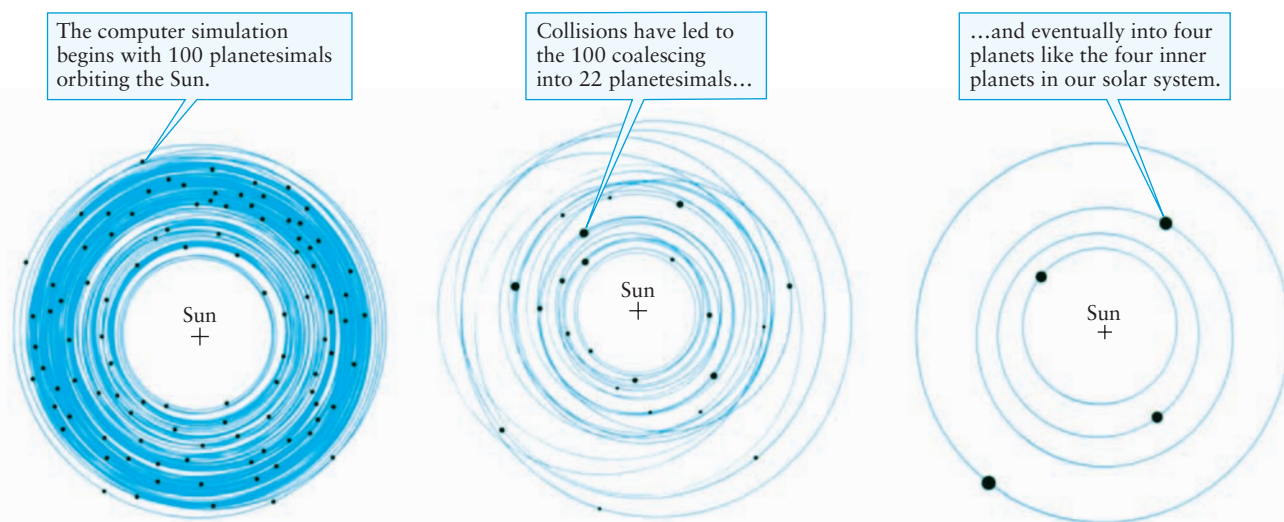


FIGURE 4-5 Accretion of the Inner Planets This computer simulation shows the formation of the inner planets as a result of myriad collisions.

not acquire coatings of hydrogen and helium like those still found on the giant planets today. However, enough hydrogen and helium were part of the debris that formed the inner four planets that these gases became their first atmospheres. Being so light, these gases quickly drifted into space, as we explore further in Chapter 6.

Because Mercury, Venus, and Mars have similar densities as Earth (in Latin, *terra*), they are composed of similar chemicals. Consequently, all four bodies are called **terrestrial planets**.

The large quantities of water that exist on Earth today, that existed on Mars, and that are likely to have existed briefly on Venus came in part from ices in the planetesimals that created these worlds and in part from impacts of water-rich debris called comets. Thus, the Nice model shows how our world came to have so much metal and so little hydrogen and helium, as proposed earlier in the chapter.

One planetesimal impact on the young Earth was particularly notable. When the young Earth had between 80% and 90% of its present mass, it was struck by a planetesimal with a mass that may have been a few times the mass of Mars (Mars has roughly 10% of Earth's present mass). Computer simulations of such an impact indicate that it would have splashed a lot of debris off Earth's surface and into orbit and tilted Earth's rotation axis. This event occurred about 4.5 billion years ago, within 100 million years after Earth started forming. The orbiting debris created a short-lived ring, which clumped together (in a process analogous to how planetesimals grew from collisions) and formed our Moon.

DEBRIS: REMNANTS IN THE SOLAR SYSTEM

4-5 The changing orbits of the giant planets spread debris throughout the solar system

The Nice model also provides interesting insights into the evolution of the outer solar system after the four giant planets were formed. These planets were so much more massive than the surviving planetesimals with which they orbited that they were able to gravitationally fling some of this remaining debris toward the inner solar system. To do so while conserving energy and angular momentum, the planets had to spiral outward in response. Lower-mass Saturn spiraled outward faster than Jupiter. Recall from Kepler's third law (Section 2-5) that as planets move farther from the Sun, the periods of their orbits lengthen. Eventually, Saturn reached a distance where it orbited the Sun once every time Jupiter orbited twice. This situation, where the periods of two orbits are related by whole numbers, is called a *resonance* because the resulting gravitational interactions between the two bodies reinforce each other. The pro-

cess is analogous to pushing someone on a swing—if you push at the right time (that is, in resonance with the person's motion), the person will go higher and higher.

As Saturn passed through this distance, resonant tugs from Jupiter pulled Saturn forward in its orbit, giving Saturn energy that enabled it to rapidly spiral further outward. Approaching Neptune and Uranus, Saturn's gravitational attraction in turn gave these bodies extra strong tugs, which in turn forced them to spiral out much farther away from the Sun than they were originally. Recall that Neptune formed closer to the young Sun than Uranus. Because it was closer to Saturn during this transformative period, Neptune received stronger gravitational tugs from Saturn (and Jupiter) than did Uranus. As a result, Neptune was forced out past Uranus.

At that time there were still billions of planetesimals in the realm of the giant planets. The gravitational effect of the outwardly moving planets caused most of these planetesimals to change orbits. Some were sent inward, but most were hurled into orbits much farther away from the Sun than they had been. The outward-flying planetesimals spread way beyond the original disk (which had a radius of 30 AU, as noted above). Billions of them were flung into a bagel-shaped volume beyond the orbit of Neptune that is now called the **Kuiper belt** (Figure 4-6), named after Gerard Kuiper, who first proposed the existence of such a region in 1951. Many more planetesimals were sent out even further than the Kuiper belt, creating a spherical distribution of debris called the **Oort cloud**.

Kuiper belt objects, or **KBOs**, were first observed in 1992. Most of the Kuiper belt objects are ice-rich comets. Some of them have developed highly elongated orbits that occasionally bring them closer to the Sun than Earth's orbit. When this happens, the Sun's radiation vaporizes some of their ices, producing the long, flowing tails that we associate with comets.

Kuiper belt objects are distributed from about 30 AU to about 50 AU from the Sun. At least 1500 of them have been cataloged. The smallest Kuiper belt object observed so far, discovered using the Hubble Space Telescope in 2009, is only 1 km across. Astronomers believe that there are many Kuiper belt objects much smaller than this. Even nonscientists, however, are familiar with one of several large bodies that the giant planets flung out—Pluto. Pluto became locked in a 2:3 resonant orbit with Neptune, meaning that it orbits the Sun twice in the same time that Neptune orbits 3 times. The gravitational effects of resonances are so pronounced that thousands of other bodies, called *plutinos*, have the same type of resonant orbit around the Sun (two orbits of each plutino for every three orbits of Neptune) as does Pluto. Astronomers have observed several other bodies similar in size to Pluto that were also flung far into the Kuiper belt. All bodies orbiting beyond Neptune that have enough mass to pull themselves into spherical shape are called *plutoids*. Pluto is one such body.

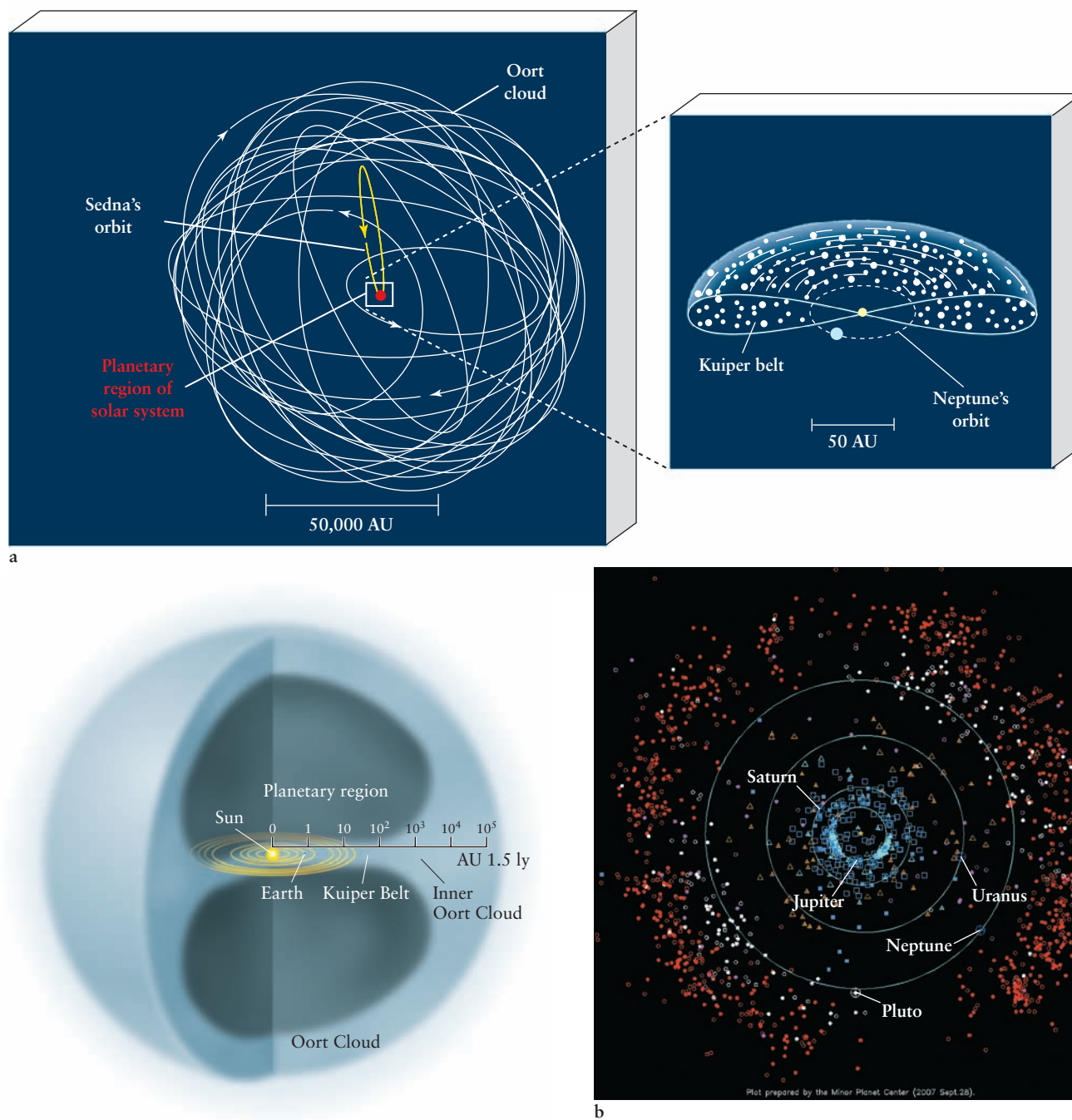


FIGURE 4-6 The Kuiper Belt and Oort Cloud (a) The classical Kuiper belt of comets spreads from Neptune out 50 AU from the Sun. Most of the estimated 200 million Kuiper belt comets are believed to orbit in or near the plane of the ecliptic. The spherical Oort cloud with a radius of perhaps 50,000 AU and containing billions of comets, extends *much farther*

out beyond the Kuiper belt than shown in this illustration. The top figures show orbits of known Oort cloud and Kuiper belt objects, while the lower-left image summarizes the locations of various regions of the solar system. (b) Positions of known bodies in the Kuiper belt and Oort cloud. (b: Gareth Williams/Minor Planet Centre.)

Focus Question 4-4

Pluto takes 248 years to orbit the Sun once. Are the orbital periods of the plutinos greater than, less than, or equal to 248 years? Why?

The Oort cloud extends out perhaps 50,000 AU, a quarter of the way to the nearest star (see Figure 4-6). Most bodies in that distribution are believed to be a few kilometers in size or smaller. One candidate Oort cloud

object, Sedna, is in an extremely elliptical orbit (Figure 4-6a), taking it between 76 and 976 AU. *All* the objects that orbit the Sun farther than Neptune are called **trans-Neptunian objects**, or **TNOs**. Plutinos and plutoids are examples of TNOs.



R I V U X G

FIGURE 4-7 An Asteroid This picture of the asteroid Gaspra was taken in 1991 by the *Galileo* spacecraft on its way to Jupiter. The asteroid measures $12 \times 20 \times 11$ km. Millions of similar chunks of rock orbit in the asteroid belt between the orbits of Mars and Jupiter. (NASA)

4-6 The asteroid belt is leftover debris

Now we turn to the major reservoir of leftover rock and metal debris remaining in the *inner* solar system, the **asteroid belt**, located between Mars and Jupiter. Orbiting in this region are millions of planetesimals and smaller pieces of debris. The total mass of debris in the asteroid belt is about 5% of our Moon's mass, much less than the mass of a planet. There were more bodies in the asteroid belt in the distant past, but impacts and the gravitational tug of Jupiter pulled some bodies away and kept the remaining debris spread out, thereby preventing it from ever coming together to form a larger body.

The largest asteroid, Ceres, has a diameter of about 900 km (compared to Earth's diameter of 12,756 km). It was the first asteroid discovered, in 1801. Today Ceres is also classified as a dwarf planet, as we will discuss shortly. The next largest, Vesta and Pallas, are each about 500 km in diameter. Still smaller asteroids are increasingly numerous. Many thousands of kilometer-sized asteroids have been observed. A close-up picture of the asteroid Gaspra is shown in **Figure 4-7**. Some asteroids have their own moons, such as the asteroid Ida, which is orbited by heavily cratered Dactyl.

Some asteroids in the inner solar system orbit between the asteroid belt and the Sun. Many of these objects have highly elliptical orbits that take them across the paths of some of the planets, including Earth.

4-7 The infalling debris from the giant planets led to the Late Heavy Bombardment

Earth, along with every other body that orbits the Sun, has been struck by countless pieces of space debris throughout its life. Indeed, these impacts still occur today. Depending on the speed and angle at which they strike land, impacting bodies larger than about 0.5 m create scars, called **craters**. Most of the craters on our planet's surface have been erased, as we will study in Chapter 6, but we can still see numerous ones on our Moon, as well as on nearly all the other bodies with solid surfaces in the solar system (**Figure 4-8**). Venus is the most notable exception, as it is periodically resurfaced.

During the first 400 million years that the planets existed, the rate of impacts declined steadily. But then, about 4.1 billion years ago, the impact rate skyrocketed as planetesimals in the asteroid belt were knocked out of orbit as debris flung inward by the giant planets (see Section 4-5) struck them. Numerous asteroids were pushed into orbits that crossed the paths of the inner planets. Thus began the *Late Heavy Bombardment*, which ended as these errant asteroids were destroyed by either striking something, falling into the Sun, or eventually being flung out of the inner solar system.



R I V U X G

FIGURE 4-8 Our Moon This photograph taken by astronauts in 1972 shows part of the near side and part of the more heavily cratered far side of our Moon. Its numerous craters were produced by impacts of leftover rocky debris from the formation of the solar system. Age-dating of lunar rocks brought back by the astronauts indicates that the Moon is about 4.5 billion years old. Most of the lunar craters were formed during the Moon's first 700 million years of existence, when the rate of bombardment was much greater than it is now. The large dark regions are the maria. (NASA)

The Moon was struck by so many objects during the Late Heavy Bombardment that large regions of its surface were completely obliterated, creating gigantic craters, which remained on the surface for hundreds of millions of years. These craters were eventually filled in when molten rock from inside the Moon (magma) leaked out through the thin crust below the craters, thereby forming the dark plains called maria that we see today (Figure 4-8). Mercury, which is particularly dense when compared to the other inner planets, may have lost much of its lower-density outer layers in a collision that occurred during this time.

The Moon's virtual lack of atmosphere has helped preserve important information about the early history of the solar system. Radioactive dating of Moon rocks brought back by the Apollo astronauts indicates that the rate of impacts declined dramatically about 3.8 billion years ago. That time is taken to be the end of the Late Heavy Bombardment. Since that time, impact cratering has continued but at a very low level.

CATEGORIES OF THE PRESENT-DAY SOLAR SYSTEM

4-8 The categories of solar system objects have evolved

While it makes sense that the objects in the asteroid belt are called asteroids, the naming of bodies there and throughout the solar system is more complicated than that. Traditionally, an **asteroid** is a piece of space debris composed primarily of rock and metal that is larger than about 10 m (no exact minimum size has been agreed upon). Space debris smaller than this and composed primarily of rock and metal are called **meteoroids**. All debris that is a roughly equal mix of ice and rock are called **comets**. More precisely, the solid body of ice and rock that exists even when a comet has no tail is called a *comet nucleus*. The asteroid belt contains all three types of debris. Indeed, all the remaining planetesimals have become today's asteroids, meteoroids, and comets.

Another layer of terminology about bodies orbiting the Sun (and, by inference, any other stars) was introduced in 2006, when the International Astronomical Union redefined the term *planet*. According to the most recently accepted definition, a **planet** is a celestial body that (1) orbits the Sun or another star; (2) has enough

mass so that its own gravitational attraction causes it to be essentially spherical, rather than being a large, irregular rock in space; and (3) has enough gravitational attraction to clear its neighborhood of other orbiting debris. Planets “clear their neighborhoods”

by either pulling surrounding debris onto themselves or, in some cases, flinging debris far away from their orbits. There are thus eight planets in the solar system: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune.

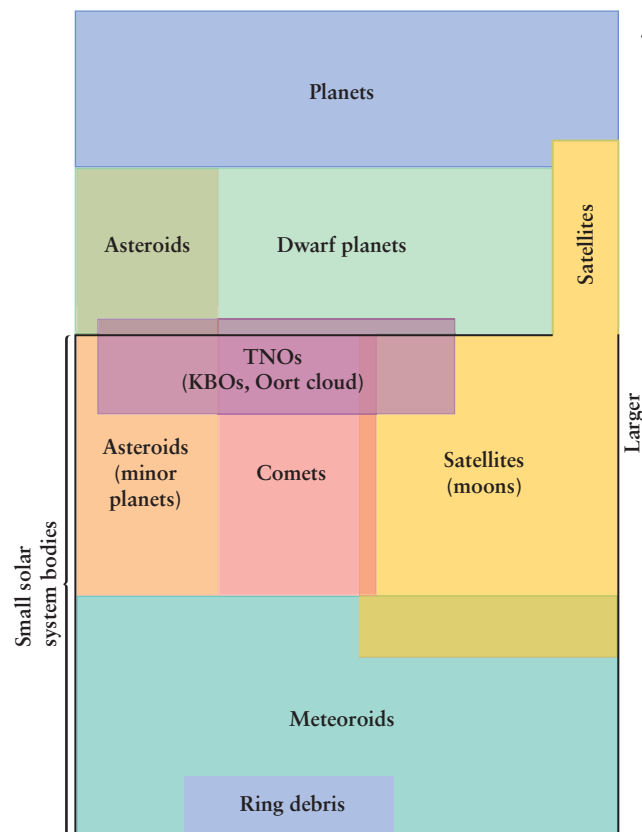


FIGURE 4-9 Different Classifications of Solar System Objects Some of the definitions of the different types of objects in the solar system overlap. For example, the largest asteroids are also classified as dwarf planets; various trans-Neptunian objects (TNOs) are asteroids or comets; some comets are satellites of Jupiter; some Kuiper belt objects (KBOs) are satellites of other KBOs. Furthermore, TNOs exist in two groups: KBOs and Oort cloud bodies. Some moons are as large as dwarf planets. Indeed, some moons are as large as small planets!

A **dwarf planet** satisfies conditions (1) and (2) but not condition (3). These objects are smaller than planets with less mass and, hence, less gravitational attraction. This definition is the reason that Pluto was demoted from planet status. So far, of the myriad candidate bodies, astronomers have agreed on five dwarf planets: Pluto, Ceres, Haumea, Makemake (pronounced *mah-kay mah-kay*), and Eris.

A **moon** or **natural satellite** is an object orbiting a larger body that is, in turn, orbiting the Sun or another star. *Rings* of billions of dust particles and mountain-sized pieces of debris in orbit around Jupiter, Saturn, Uranus, and Neptune are yet another category of solar system object.

Everything in the solar system that is not a planet, dwarf planet, moon, or part of a ring system, is called a **small solar system body (SSSB)**. The tricky part is that the largest asteroids are dwarf planets, whereas the vast majority of asteroids, as well as all meteoroids and all comets, are small solar system bodies. **Figure 4-9** summarizes these relationships.

Focus Question 4-5

How might some of the meteoroids in the solar system come from asteroids?

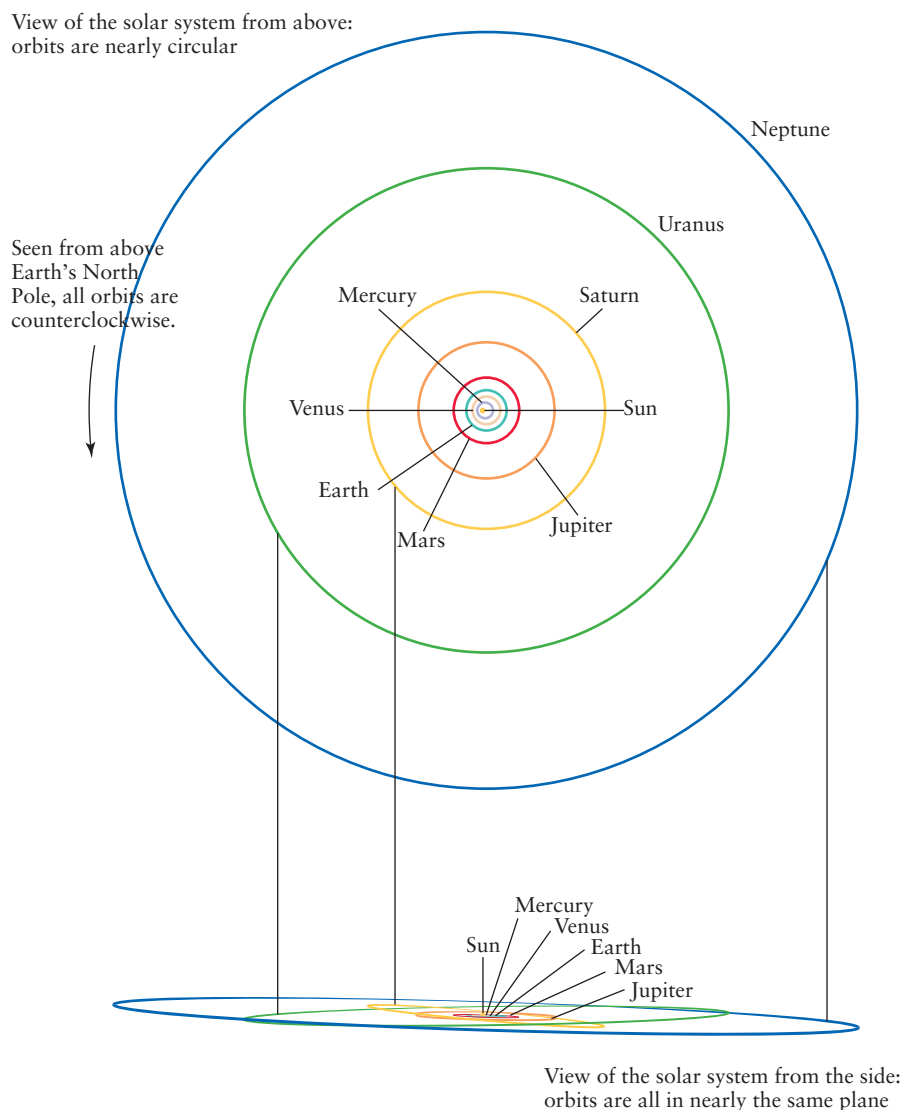


FIGURE 4-10 The Solar System This scale drawing shows the distribution of planetary orbits around the Sun. All orbits are counterclockwise because the view is from above Earth's North Pole. The four terrestrial planets are located close to the

Sun; the four giant planets orbit at much greater distances. Seen from above the disk of the solar system, most of the orbits appear nearly circular. Mercury has the most elliptical orbit of any planet.

4-9 The orbits of the planets are related

Because the planets formed in a disk, they should all orbit in more or less the same plane as Earth (that is, nearly in the plane of the ecliptic). Observations bear out this conclusion. The angles of the orbital planes of the other planets with respect to the ecliptic are called their **orbital inclinations**, and all are 7° or less (**Figure 4-10**).

Furthermore, according to the Nice model, all of the planets formed from the same swirling cloud of gas and dust, and hence they should all orbit the Sun in the same direction, which they do. Moreover, since the Sun was formed from the same swirling gas and dust as the planets and other members of the solar system, it should be rotating, and the planets should be orbiting in the same direction as the Sun rotates. Imagine traveling far

above Earth's North Pole and looking down at the Sun and planets. You would see the Sun rotating counterclockwise and all the planets orbiting counterclockwise, consistent with the Nice model of how the solar system formed.

3 To find out more about the solar system's origins, astronomers study the leftover interplanetary debris: asteroids, meteoroids, and comets. Many of these bodies are thought to be virtually unchanged remnants of the solar system's formation. According to radioactive dating of the oldest debris ever discovered from space (see Appendix Q: Radioactivity and the Ages of Objects), the solar system was roughly in the form we know today just under 4.6 billion years ago. As we will see throughout the book, a lot happened between the formation of the

universe some 13.8 billion years ago and the formation of the solar system, some 9 billion years later.

On the young Earth, one of the conditions that allowed for the beginning of biological evolution was the presence of liquid water. Strong evidence indicates that liquid water existed on Mars and may still exist inside it and inside Jupiter's moons Europa and Ganymede, as well as Saturn's moons Titan and Enceladus, raising the hotly debated question of whether simple life has evolved on those worlds as well.

4-10 The Sun developed while the planets matured

During the 100 million years that the inner planets were completing their formation, the protosun was also evolving. Throughout this time, the temperature and pressure at the center of the slowly contracting protosun continued to increase. Finally, its core reached about

10^7 K, which is hot enough to initiate hydrogen fusion, in which hydrogen is converted into helium. Such fusion generates energy in the form of gamma rays, which, slamming into the gases in the protosun, created enough pressure inside the protosun to stabilize it, here meaning preventing it from expanding or contracting. At that point, the Sun became a star. As noted earlier, some of the energy generated in the Sun's core leaks out, which is why the Sun shines. Hydrogen fusion continues in the Sun's core today and the energy generated there over the past 4.5 billion years has maintained the Sun's size and temperature, and thereby enabled Earth to sustain life.

Sunlike stars take approximately 100 million years to form from a nebula and settle down, which means that the Sun probably became a full-fledged star around the time the accretion of the inner planets was completed. Radiation from the Sun heated and dispersed the remaining gases in the solar system, thereby helping to limit the sizes of the planets.

SUMMARY OF KEY IDEAS

Formation of the Solar System

- Hydrogen, helium, and traces of lithium, the three lightest elements, were formed shortly after the formation of the universe. The heavier elements were produced much later by stars and are cast into space when stars die. By mass, 98% of the observed matter in the universe is hydrogen and helium.
- The solar system formed 4.6 billion years ago from a swirling, disk-shaped cloud of gas, ice, and dust, called the solar nebula.
- The planets and other debris in the solar system today formed from gas, ice, and dust in the solar nebula orbiting the protosun.
- The outer solar system, beyond the snow line, had both dust and ice (including hydrogen and helium), while inside the snow line, such ices were vaporized by the protosun.
- Jupiter and Saturn were initially worlds of rock and metal that pulled onto themselves large amounts of hydrogen and helium, along with some water.
- Uranus and Neptune were also initially worlds of rock and metal, but they attracted more water and less hydrogen and helium than the other giant planets.
- The Nice model of solar system formation proposes that in the outer solar system Jupiter formed first, followed by Saturn, and then by Neptune and Uranus, which were then flung out to their present orbits by gravitational forces from Jupiter and Saturn.
- The four inner planets formed through the collisions of Moon-sized bodies, probably after the outer four planets were formed.

- The Sun formed at the center of the solar nebula. After about 100 million years, the temperature at the protosun's center was high enough to ignite thermonuclear fusion reactions.
- For 800 million years after the Sun formed, impacts of asteroidlike objects on the young planets dominated the history of the solar system.

Categories of Solar System Objects

- Astronomical objects smaller than the eight planets are classified as dwarf planets or small solar system bodies (SSSBs).
- A variety of other names, including asteroids, comets, meteoroids, trans-Neptunian objects, plutinos, plutoids, Kuiper belt objects (KBOs), and Oort cloud objects, overlap with the designations “dwarf planet” and “SSSB.”
- KBOs and Oort cloud objects are trans-Neptunian objects—they orbit farther from the Sun than the outermost planet.
- To date, five objects—Pluto, Ceres, Eris, Haumea, and Makemake—have been classified as dwarf planets.
- Other objects orbit the Sun beyond Neptune. At least 1500 KBOs have been observed. A few potential Oort cloud objects have also been identified.

WHAT DID YOU THINK?

- 1 *How many stars are there in the solar system?* The Sun is the only star in our solar system.
- 2 *Were the sun and planets among the first generation of objects created in the universe?* No. All matter and energy were created by the Big Bang. However, much of the material that exists in our solar system was pro-

cessed inside stars that evolved before the solar system existed. The solar system formed billions of years after the Big Bang occurred.

3 *How long has Earth existed, and how do we know this?* Earth formed along with the rest of the solar system about 4.6 billion years ago. The age is determined from the amount of radioactive decay that has occurred on Earth.

Review Questions

- About how long after the universe came into existence did our solar system form?
 - 0 years (they formed together)
 - a million years
 - 10 million years
 - a billion years
 - 10 billion years
- Pluto is most similar in composition to which of the following objects?
 - Eris
 - Jupiter
 - our Moon
 - Earth
 - the Sun
- Which planets are terrestrial and which are giant?
- What is the Nice model's explanation for why there is no planet in the vicinity of the asteroid belt?
- Which giant planet formed first?
- According to the Nice theory, where did the Kuiper belt objects and Oort cloud objects come from?

7. What created most of the craters in the solar system? What else could create craters? (*Hint:* Think of craters on Earth.)

8. Jeans instability is responsible for what event in the evolution of the solar system?

9. Name and briefly describe one small solar system body.

10. Name one dwarf planet (other than Pluto) and state where it is located.

11. Asteroids, meteoroids, and comets are also classified as what (one) kind of object today?

Advanced Questions

12. Why would it have been hard for the first generation of stars to have planets? If they did have planets, of what would those bodies have been composed?

13. If Saturn had not formed, how would Jupiter have evolved after entering inside the snow line, and how would Jupiter have then affected the formation and evolution of Earth?

14. What is significant about the snow line described in this chapter?

Got It?

15. Which elements on Earth were created during the Big Bang?

16. How long has the universe existed and how long has the solar system existed?

This page intentionally left blank



CHAPTER 5

R I V I U X I G The Orion Nebula, 1350 ly away, as imaged by the Hubble Space Telescope. Part of the much larger Orion Molecular Cloud, the Orion Nebula contains more than 3000 stars and many more in the process of forming. The bright central region is the home of the Trapezium, four young, bright stars that cause the surrounding gas and dust to glow. The bright region in the upper left, denoted M43, is glowing due to the radiation from one young, massive star. (NASA, ESA, M. Robberto [STScI/ESA] HST Orion Treasury Project Team)

Exoplanets

WHAT DO YOU THINK?

- 1 Have any planets been observed orbiting other stars?
- 2 Have any Earthlike planets been discovered orbiting in the habitable zones around other stars?
- 3 Have astronomers discovered evidence for life on any planets orbiting other stars?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

In this chapter you will discover

- how astronomers detect exoplanets
- that newly forming stars and planetary systems are being discovered every year
- that disks of gas and dust, as well as planets, are being observed around a growing number of stars
- planets orbiting other stars have sizes ranging at least from 0.2 to 25 Earth radii
- information about exoplanet atmospheres is beginning to emerge
- that most stars have planets orbiting them
- where astronomers are looking for extraterrestrial life

EXOPLANETS—PLANETS OUTSIDE OUR SOLAR SYSTEM

Understanding the process by which the solar system formed is being aided by observations of different stages of planet formation occurring around other young stars. Although astronomers have been searching for planets outside the solar system for centuries, it was only in the 1980s that observations leading to their discovery were first made.

5-1 Protoplanetary disks are a common part of the star-forming process

Stellar evolution and star formation are ongoing processes. You can see a cloud in which stars are forming today with your naked eye, namely, the middle “star” in Orion’s sword (see the figure opening this chapter). This isn’t a star at all! Called the Orion Nebula or Great Nebula of Orion, it is actually part of a gigantic interstellar cloud region called the Orion Molecular Cloud in which new stars and planets are coming into existence.

Detecting and studying newly forming stars and planets require observations in various parts of the electromagnetic spectrum outside the visible. Recall from Chapter 3 that visible light provides limited astronomical information; you can see in Figure 3-32c and Figure 4-2b that infrared images reveal even more of the matter distributed in this region. Likewise, ultraviolet images provide insights into star and planet formation unavailable from visible images. Observing with telescopes sensitive to infrared and ultraviolet wavelengths, we can see the newly forming stars and planets deep inside obscuring clouds of gas and dust. Furthermore, radio observations provide information about stellar rotation and the properties of magnetic fields surrounding planets outside the solar system.

As discussed in Section 4-2, astronomers have discovered stars in early stages of formation. Most of these protostars (conceptually analogous to the protosun) are surrounded by protoplanetary disks. These disks are similar to what the solar nebula was believed to have been like more than 4.5 billion years ago.

Along with disks of gas and dust, large numbers of comets and asteroids have been observed orbiting a few

stars. The first system discovered to have comets was the young star Beta Pictoris. The comets were detected as a result of the gas they emitted when passing close to their stars.

The first star observed to have asteroids in orbit around it is Zeta Leporis, a young star just 70 ly from Earth. Its asteroid belt, located about the same distance from that star as our asteroid belt is from the Sun, contains an estimated mass 200 times greater than the mass of our asteroid belt. Although pieces of asteroid debris have not been directly observed, their presence has been inferred from the existence of hot (room-temperature) dust surrounding Zeta Leporis. The most likely explanation for this high-temperature dust is that it is being generated by the collisions of asteroids. As noted in Section 4-6, our solar system has one asteroid belt, located between Mars and Jupiter. Observations reveal that the star Epsilon Eridani has two asteroid belts, along with planets and a ring of icy debris. In 2013, astronomers discovered an asteroid belt around the bright star Vega, in the constellation Lyra.

Although astronomers had expected to find disks of gas and dust orbiting *young* stars, consistent with the theory explaining the formation of the solar system, they have also found disks of dusty debris orbiting stars as old as the Sun (so they are over 4 billion years old). These stars are all known to have planets, and the rubble surrounding them was most likely formed by the collisions of asteroids and comets with each other and perhaps with the planets in these systems.

5-2 Astronomers have at least seven different ways of detecting planets outside our solar system

1 Planets were first observed orbiting other stars in 1992. Astronomers use a variety of different methods to detect these planets. Each method provides different, and sometimes complimentary, information about the planets and the stars they orbit.

Distorted protoplanetary disks Indirect observational evidence for the existence of planets outside the solar system, called **exoplanets** or *extrasolar planets*, first came from the observed distortions of protoplanetary disks. If one or more planets orbit in a protoplanetary disk around a young star, their gravitational pull will

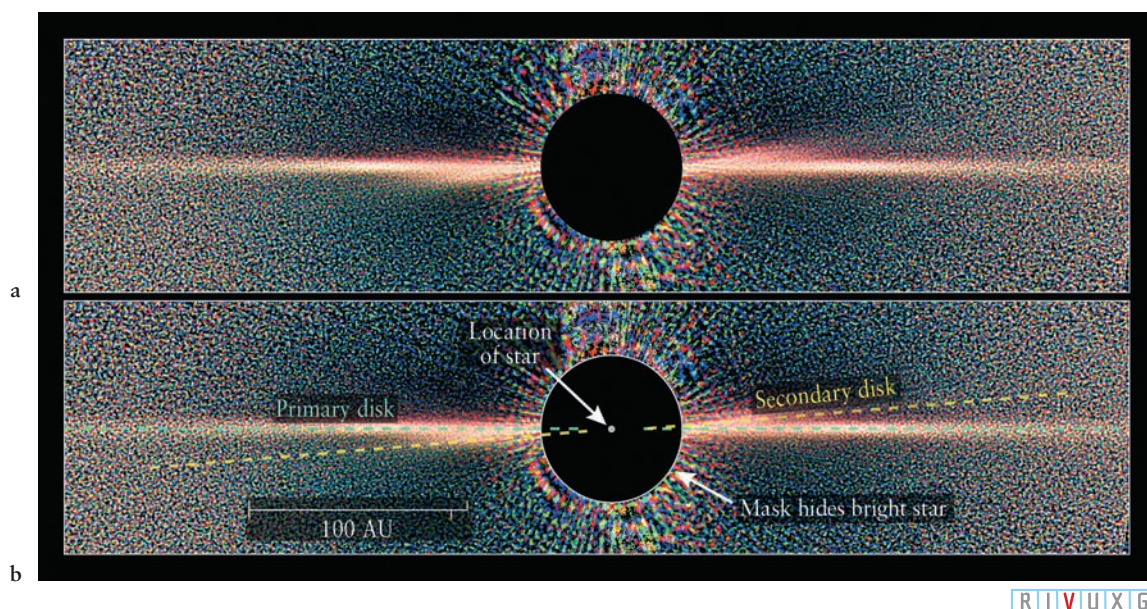


FIGURE 5-1 A Circumstellar Disk of Matter (a) Hubble view of Beta Pictoris, an edge-on disk of material 225 billion km (140 billion mi) across that orbits the star Beta Pictoris (blocked out in this image) 50 ly from Earth. Twenty million years old, this disk is believed to be composed primarily of iceberlike

bodies that orbit the star. The smaller disk is believed to have been formed by the gravitational pull of a roughly Jupiter-mass planet in that orbit. Because the secondary disk is so dim, the labeling for this image is added in (b). (NASA/ESA/D. Golimawski)

affect the disk of gas and dust around the star, causing the disk to clump or warp or become off-center from its star. An example of serious warping can be seen in the edge-on disk surrounding the star Beta Pictoris. This star and the material that orbits it formed only 20 million years ago. In 2006, astronomers discovered that Beta Pictoris actually has two disks (**Figure 5-1**), tilted slightly with respect to each other. Computer models show that a Jupiterlike planet orbiting out of the plane of the big disk would attract debris from that disk, thereby forming the smaller disk in the plane of its orbit. As noted above, Beta Pictoris also has millions of comet nuclei, which were discovered in 2001. By studying such systems as Beta Pictoris, we can also begin addressing such questions as whether Earth may have acquired its water from comets in the young solar system.

This interaction between planets and disks is seen in other systems. The nearby star Fomalhaut has an off-centered disk of gas and dust. This star and its entourage are only some 200 million years old. Observations reveal that this distortion is caused by one or more planets tugging on the gas as they orbit.

Transit photometry method Whereas observing distorted protoplanetary disks was one of the first methods of discovering exoplanets, the most successful method has been to look for exoplanets in older systems that block some of their star's light as the planet transits between its star and Earth. In these systems, the disks have already dissipated. Because most planets are not located between us and the stars they orbit for very

long, using this technique requires watching a fixed set of stars continuously over a period of years for brightness variations caused by transits. NASA's Kepler space telescope, which operated between 2009 and 2013, is to date the most successful collector of transit photometry; it stared continuously at over 100,000 stars, continually recording their brightnesses. When planets passed between one of these stars and us, the star's brightness dipped, as shown in **Figure 5-2a**.

By the time it failed, Kepler had identified 3846 candidate planets by the light curves of their stars. Some of these variations in intensity were due to pairs of orbiting stars eclipsing each other (rather than the passage of a planet), high-energy particles striking Kepler's electronics, or other electronics issues. Each candidate observation must be carefully analyzed. As of the publication of this text, 978 exoplanets were confirmed to have been discovered by Kepler.

Radial velocity method The second most common method of detecting exoplanets uses the fact that the expression "planets orbit stars" is only an approximation to the actual activity of star-planet systems. In fact, all pairs of orbiting objects orbit around a common point, called their center of mass, located directly between them (**Figure 5-2b**). Earth and the Moon, for example, orbit around a point located 1712 km (1064 mi) beneath Earth's surface. When one of the bodies is much more massive than the other, like Earth and the Sun, the center of mass is very close to the center of the more massive body. This leads to the intuitive expression: "Earth orbits the Sun."

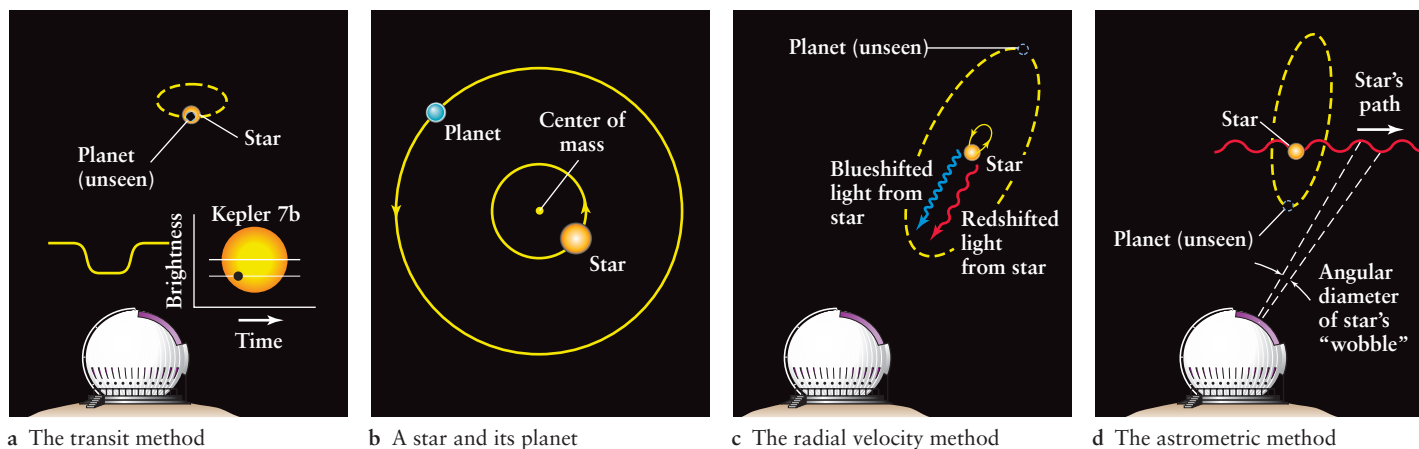


FIGURE 5-2 Three Methods of Detecting Exoplanets (a) If a planet happens to move in a plane that takes it across its star (that is, the planet transits the star), as seen from Earth, then the planet will hide some of the starlight, causing the star to dim from our perspective. This change in brightness will occur periodically and can reveal the presence of a planet. (b) A planet and its star both orbit around their common center of mass, always staying on opposite sides of that point. The star's motion around the center of

mass often provides astronomers with the information that a planet is present. (c) As a planet moves toward or away from us, its star moves in the opposite direction. Using spectroscopy, we can measure the Doppler shift of the star's spectrum, which reveals the effects of the unseen planet or planets. (d) If a star and its planet are moving across the sky, the motion of the planet causes the star to orbit its center of mass. This motion appears as a wobbling of the star across the celestial sphere.

When a massive exoplanet is orbiting close to its star, their center of mass is not near the star's center. As a result, if the planet has any motion toward and then away from us in its orbit, then so, too, does its star. We can detect this motion through the Doppler shift of the star's spectrum (as presented in Section 3-2). The length of time of one cycle of Doppler shifts (toward us, then away from us, then back to start) is the period of the planet's orbit. Modern spectroscopic techniques are so good that we can detect stars approaching or receding at speeds as low as 4 km/h (2.5 mi/h). Because the motion of the star toward or away from us is called its radial motion, this technique (Figure 5-2c) is commonly called the *radial velocity method* of detecting planets. The first exoplanet orbiting a Sunlike star fusing hydrogen into helium in its core was discovered in 1995 orbiting the star labeled 51 Pegasi (51 Peg) in the constellation Pegasus.

Astrometric method The radial velocity method used the fact that many planets can make their stars periodically move toward and then away from us. Many planets can also make their stars appear to wobble on the celestial sphere. Many planets orbit their stars with a component of their motion perpendicular to our line of sight—that is, some of their motion is on the celestial sphere. This is analogous to the motion you would see if you put your arm straight out in front of you and moved it side to side. The stars orbiting with these planets move around their common centers of mass with motion that appears to be moving back and forth on the celestial sphere. In 2009, the Hubble Space Telescope imaged the wobble of the star Gliese 876, revealing the presence of a previously

undetected planet orbiting it. Several planets had been detected orbiting it previously using the radial velocity method. The astrometric method is shown in Figure 5-2d.

Direct imaging method We cannot yet resolve most of the planets we have detected. Direct imaging is extremely challenging because even planets like Jupiter, which scatter most of the light that strikes them, are very dim compared to the stars they orbit. As mentioned in Chapter 4, in order to be classified as a star, an object has to have sufficient mass to compress its insides enough to force fusion to occur in its core. The lowest mass for which this can happen is about $0.08 M_{\odot}$ (solar masses), which is equal to about $75 M_{\text{Jupiter}}$. Therefore, it is perhaps not surprising that the first planetlike body orbiting another object outside the solar system, directly imaged in 2004, was one in which the second object, called a brown dwarf labeled 2M1207, is not quite a star; it has only $0.025 M_{\odot}$, so no fusion occurs in it. A brown dwarf is much dimmer than a star so that it does not overwhelm the light from the planetlike body (Figure 5-3).

The smaller companion to 2M1207, denoted 2M1207b, has a mass between 5 and 8 times the mass of Jupiter. Because the larger body does not shine brightly, the dimmer, planetlike body is easily visible. However, because the larger body is not a star, the smaller one is not considered a planet.

Combining observations of disks of gas and dust around young stars and direct imaging, the planet Fomalhaut b was first observed in 2004 (Figure 5-4). This exoplanet has an extremely eccentric orbit (much more so than any planet in our solar system). Its orbit ranges

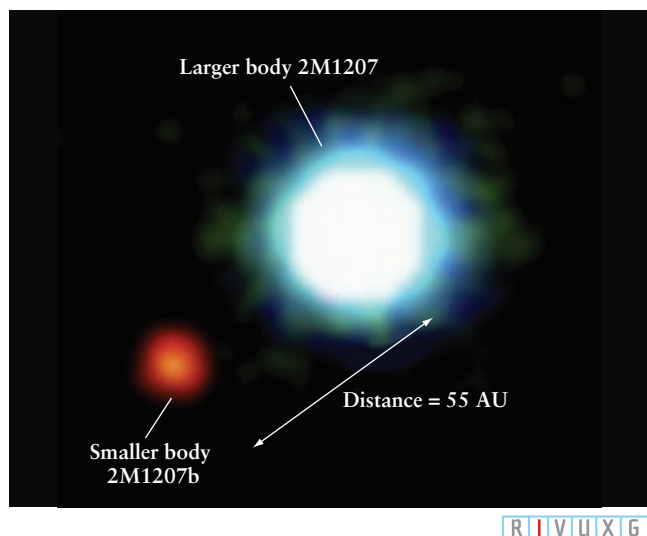


FIGURE 5-3 Image of an Almost Extrasolar Planet This infrared image, taken at the European Southern Observatory, shows the two bodies 2M1207 and 2M1207b. Neither is quite large enough nor massive enough to be a star, and evidence suggests that 2M1207b did not form from a disk of gas and dust surrounding the larger body; hence, it is not a planet. This system is about 170 ly from our solar system in the constellation Hydra. (ESO/VLT/NACO)

from 5 to 29 times further from its star than Saturn is from the Sun. The pull of the planet (and possibly others) is so great that the center of Fomalhaut's disk is displaced 15 AU from the star (1 AU is the average distance from Earth to the Sun). Furthermore, it appears Fomalhaut b will enter the disk in 2032, causing great havoc as it pulls some of the orbiting debris onto itself and flinging other debris out of orbit.

Besides Fomalhaut b, 16 other planets have been resolved. These are all large planets that are bright in the light scattered from their stars (**Figure 5-5**).

Microensing method In 2004, astronomers began finding exoplanets by using a property of space that Albert Einstein discovered. His theory of general relativity, which we explore in detail in Chapter 12, shows that matter warps its surrounding space, causing, among other effects, passing light to change direction. This phenomenon is quite analogous to how light passing through a lens changes direction and is focused—light can also be focused by gravity, an effect called **microlensing** (or *gravitational microlensing*). **Figure 5-6** shows how a star with a planet, passing between Earth and a distant star, can focus the light from the distant star, causing it to appear to change brightness. This change occurs twice, once as the distant star's light is focused toward us by the closer star and again when the light is focused toward us by the planet and the star together.

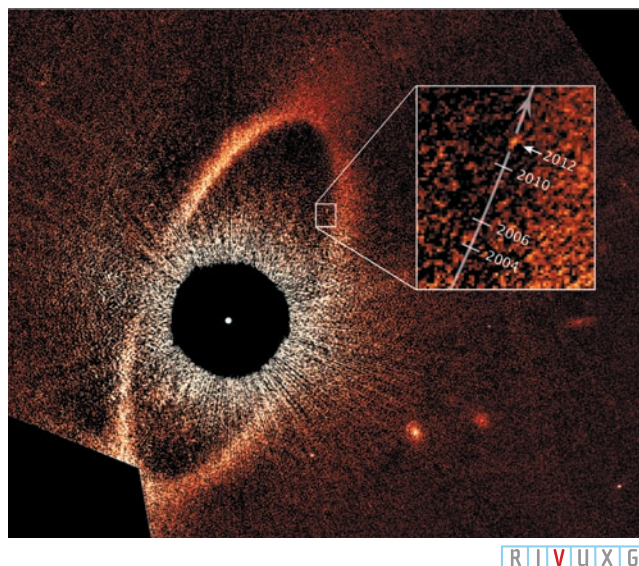


FIGURE 5-4 Visible Image of an Exoplanet The star Fomalhaut, blocked out so that its light does not obscure the disk, is surrounded by gas and dust in a ring whose center is separated from the star by 15 AU, nearly as far as Uranus is from the Sun. This offset is due to the gravitational effects of giant planet Fomalhaut b orbiting the star. This system is 25 ly from Earth. The dimmer debris in that system and between it and Earth scatters light that is called “noise” in this image. (NASA, ESA, and P. Kalas [University of California, Berkeley and SETI Institute])

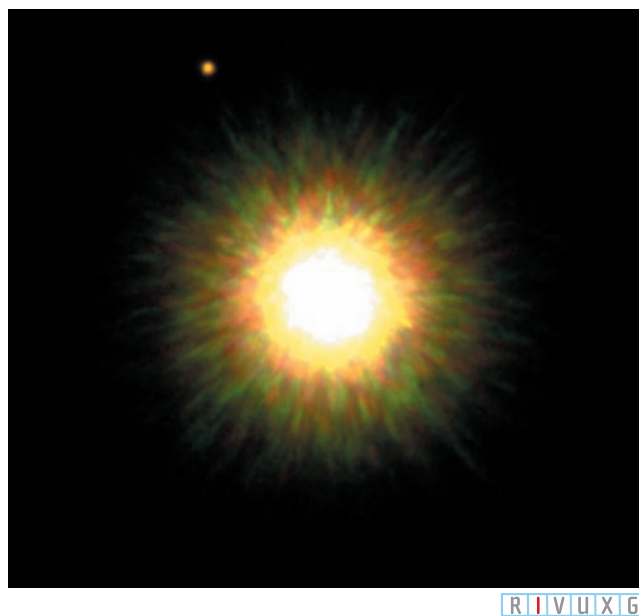
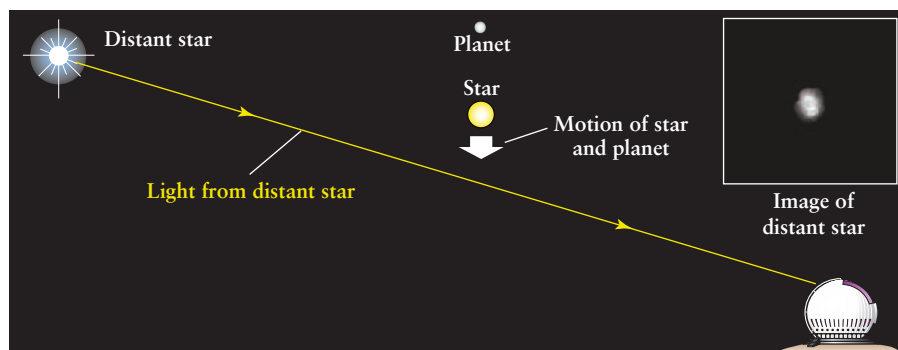
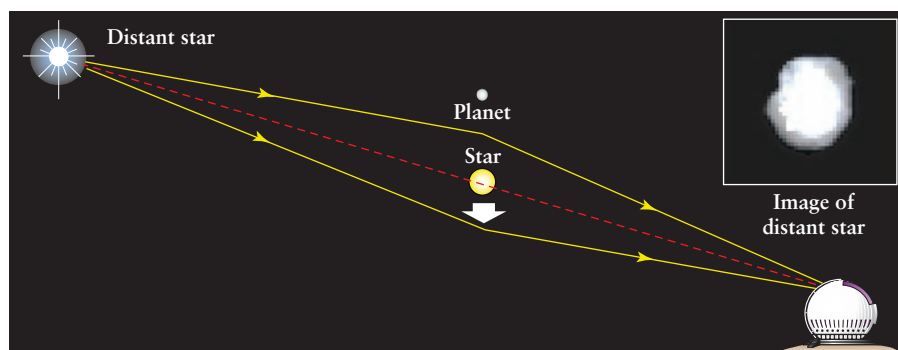


FIGURE 5-5 Direct Image of an Extrasolar Planet A planet with 8 times the mass of Jupiter orbiting the Sunlike star 1RXS 1609. (Gemini Observatory, D. Lafreniere, R. Jayawardhana, M. van Kerkwijk [University of Toronto])

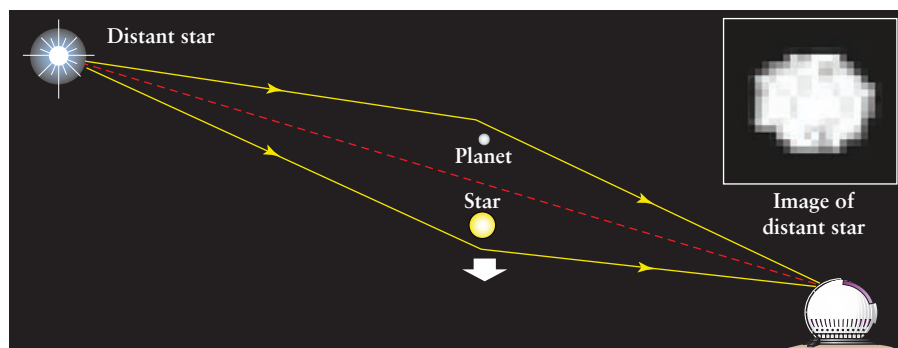
Pulsar timing method One of the rarest methods for detecting planets is their influence on the remnants of stars. All stars evolve, meaning that their chemistry, size, temperature, and other properties all change with time.



a No microlensing



b Microlensing by star



c Microlensing by star and planet

FIGURE 5-6 Microlensing Reveals an Exoplanet (a) Gravitational fields cause light to change direction. As a star with a planet passes between Earth and a more distant star (b), the light from the distant star is focused toward us, making the distant star appear brighter. The focusing of the distant star's light occurs twice, once by the closer star and once by the closer star and its planet (c), making the distant star change brightness. For these simulations, the closer star and planet are 17,000 ly away, while the distant star is 24,000 ly away.

Eventually they stop fusing and either explode or simply die out. Stars with between 8 and 25 $M_{\odot}$ explode as powerful supernovae that leave cores of between 1.4 and 3 $M_{\odot}$ composed nearly entirely of neutrons. Many of these *neutron stars* are rotating hundreds or thousands of times a second and, like the Sun, they have magnetic fields that whip around with the rotation of the neutron star. These whirling magnetic fields generate pulses of radio waves, visible light, and other radia-

tion; hence, they are called *pulsars*. If we are in the line of sight of the pulses, then we see them. Pulsars emit among the most regular pulses that we have ever detected, far more precisely than, for example, the most accurate wristwatch.

If a pulsar has planets orbiting it, then the motion of the pulsar around the center of mass of the planet(s)–pulsar system can be detected in the changes in the timing of its pulses. Despite the fact that pulsars are extremely rare compared to stars that are still shining by fusion, the first exoplanets ever discovered were found because of their effects on pulsar timing. Two planets were detected orbiting the pulsar PSR B1257+12 in 1992 and a third was detected in its orbit in 1994. It is worth noting that other methods of detecting exoplanets are being developed.

5-3 Exoplanets orbit a breathtaking variety of stars

As we have just seen, planetlike bodies have been observed around bodies that have too little mass to be classified as stars, while planets have been observed orbiting the neutron star remnants of stars that were some 20 $M_{\odot}$ before they exploded. Indeed the more than 1800 exoplanets that have been observed in orbit around stars of masses ranging from 0.08 $M_{\odot}$ to over 20 $M_{\odot}$.

While some planets are orbiting stars similar to the Sun, many planets orbit burned-out remnants of stars and even orbit pairs or trios of stars. Most of the planets we have found are in orbit around stars located in the disk of the Milky Way Galaxy (which is where our solar system resides), but a few have been found in large clumps of stars outside our Galaxy's disk. (The word *galaxy* when used alone is capitalized only when referring to our Milky Way.) These clumps, called *globular clusters*, contain very old stars that formed shortly after the universe began. Because it is in a globular cluster, the oldest-known planet dates back to the cluster's formation 13 billion years ago.

The closest-known exoplanet is orbiting the star Alpha Centauri B, just 4.3 light-years away. The most distant exoplanet known is located nearly 30,000 ly away.

5-4 Exoplanets with a wide range of sizes, masses, and compositions have been observed

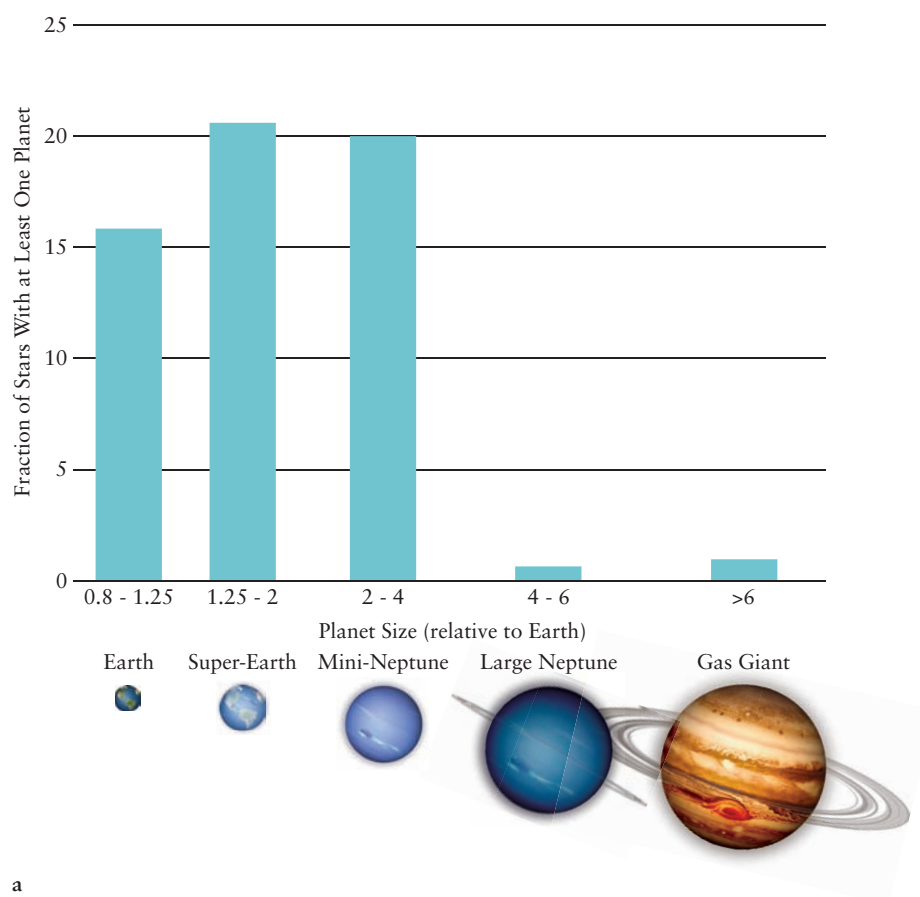
Planets of all sizes ranging from 0.2 to 25 times Earth's diameter have been discovered. This is particularly interesting because the radii of planets in our solar system don't have this range. Expanding the Nice model

to explain the masses, sizes, and locations of exoplanets is still very much a work in progress. The smallest exoplanet discovered so far, Kepler 37b, has a diameter only slightly larger than that of our Moon. It is one of at least three planets orbiting the star, Kepler 37, which has a mass of $0.8 M_{\odot}$.

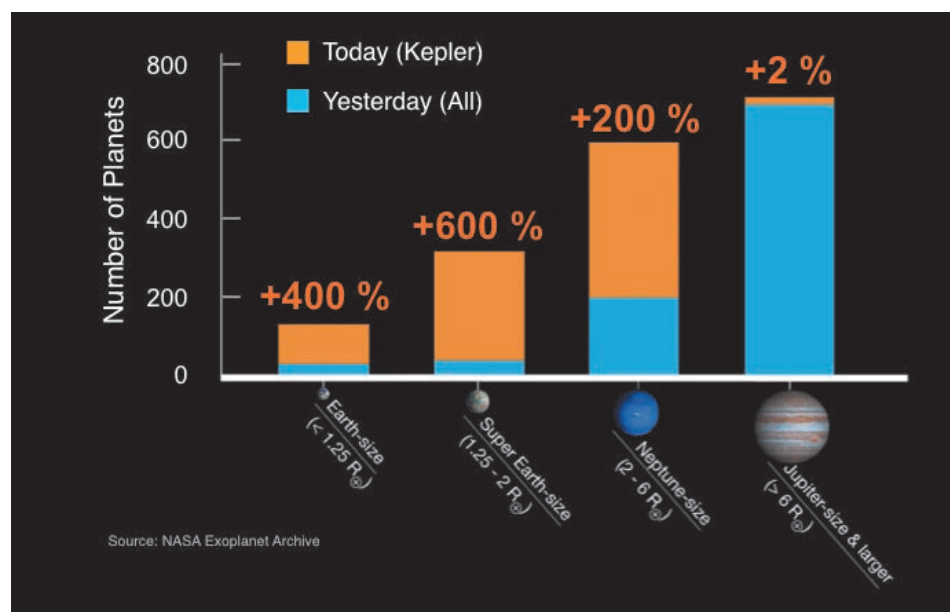
By observing transits, astronomers can also study the spectrum of a planet's atmosphere, from which we learn about that atmosphere's chemical composition. When the gas giant planet labeled HD 209458b passed in front of the star HD 209458 (Henry Draper catalog star number 209458) some 150 ly from Earth, hydrogen, sodium, oxygen, carbon, and carbon monoxide in the planet's atmosphere absorbed certain wavelengths of starlight. Therefore, the star's spectrum has extra absorption lines (introduced in Section 3-13). The Hubble Space Telescope further revealed that the outer layers of this planet are being heated so much that at least 9×10^6 kg (10,000 tons) of hydrogen are evaporating into space from it every second. In 2010, a tail created by this departing gas was observed. Based on this rate of mass loss, the planet has lost at least 0.1% of its mass over its lifetime of 5 billion years. Furthermore, observations in 2010 revealed winds in the outer cloud layers of this planet reaching 7000 km/h (4300 mi/h), more than 3 times faster than any winds known to occur in our solar system. The atmospheres of other exoplanets also include methane, carbon dioxide, and water vapor.

The transit method has also revealed that some exoplanets have much more high-density material in their cores than do our giant planets. The changing intensity of light from the star HD 149026 showed that the planet passing in front of it, which has a mass 115 times that of Earth, is much smaller than other planets of that mass. Calculations reveal that about 70 Earth masses (roughly two-thirds) of that planet is rocky material in its core. For comparison, Jupiter's terrestrial core is about 4% of its total mass.

In 2005, microlensing was used to detect a terrestrial (rocky) planet labeled Gliese 876d orbiting star Gliese 876. (The Gliese catalogs list nearby stars and their known planets.) This planet has 6.8 times Earth's mass and, because it is less than twice Earth's radius, it is classified as a **super-Earth**. **Figure 5-7a** summarizes the present naming scheme for exoplanets based on their sizes. **Figure 5-7b** shows the distribution of known exoplanet masses.



a



b

FIGURE 5-7 Exoplanet Classifications (a) This chart gives the names and sizes of the exoplanets. (b) This chart gives the distribution of exoplanets as a function of their masses. (b: NASA Ames/W Stenzel)

5-5 Stars with multiple planets have been observed

As noted previously for the star Kepler 37, stars with several planets orbiting them have been observed. Direct images of multiple planets orbiting a star have been taken. However, most are detected using the transit method. To do this, astronomers watch the largest planet transit a star several consecutive times. The interval between transits is the length of the year on that planet. When there are other planets in a star system, including ones that don't transit their star as seen from Earth, their gravitational tugs on the observed planet change the length of time it takes to orbit the star. So, if the length of a year for a transiting planet changes, then astronomers can deduce that there must be other planets in that system. In 2010, star Kepler 9 was observed to have two transiting planets and at least one nontransiting planet.

Three planets, Gliese 581e, c, and b, with masses of 1.7, 5.5, and 15.8 Earth masses, respectively, have been discovered orbiting another Gliese catalog star, Gliese 581. All of these planets orbit their stars closer than Mercury orbits the Sun, and, hence, they are likely to be in synchronous orbits. This means that they rotate at the same rate that they orbit their star; the same side of each planet continually faces the star just as the same side of our Moon continually faces Earth. Most, if not all, of these planets are believed to be composed primarily of rock and metal with water on or near their surfaces. The stars they orbit are much cooler than our Sun.

As of 2014, 460 stars have been observed with two or more planets that orbit them; several are shown in **Figure 5-8**.

5-6 Many exoplanets have extraordinary orbits, as compared to those in our solar system

In 2003, astronomers used the transit method to detect a planet only 0.024 AU from its star (one-sixteenth as far from its star as Mercury is from the Sun). That planet orbits once every 28 hours and 33 minutes. In 2011, the planet Kepler 70b was observed orbiting its star once every 5.8 hours at a distance of 0.01 AU. This planet has a surface temperature of 7200 K, hotter than the surface of our Sun. It is the planet closest to its star known so far.

In 2010, observations of exoplanets transiting their stars revealed that some of the planets are orbiting in the opposite direction to which their stars rotate (or spin). This motion is intriguing because it runs counter to the predictions of the Nice model (see Section 4-2), in which planets should orbit in the same direction as the stellar rotation, as occurs here in our solar system.

Many of the more than 1800 exoplanets discovered so far have highly elliptical orbits, with eccentricities up to at least $e = 0.97$, compared to the highest orbital eccentricity of any planet in our solar system of 0.206 for Mercury. For example, Fomalhaut b, discussed in Section 5-2, is a planet with a high orbital eccentricity. Star systems with massive planets in highly eccentric orbits are unlikely to have life-sustaining Earthlike planets, because the changing location of the massive planet is likely to prevent smaller planets from staying in stable orbits.

As you can see in Figure 5-8a, many giant planets orbit surprisingly close to their stars, much closer than Earth is to the Sun. If the Nice model of how giant planets form is correct, then these planets are spiraling inward after having formed farther away from their stars than they are now.

A dramatic consequence of the inward spiraling of planets was discovered in 2001, when the remnants of at least one planet were discovered in the atmosphere of star HD 82943 that still has at least two other planets in orbit around it. This star's atmosphere contained a rare form of lithium that is found in planets but that is destroyed in stars within 30 million years after they form. The presence of this isotope, ${}^6\text{Li}$, means that at least one planet spiraled so close to the star that it was vaporized.

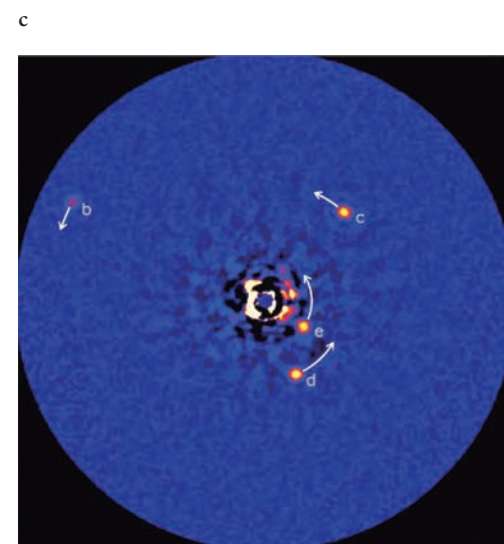
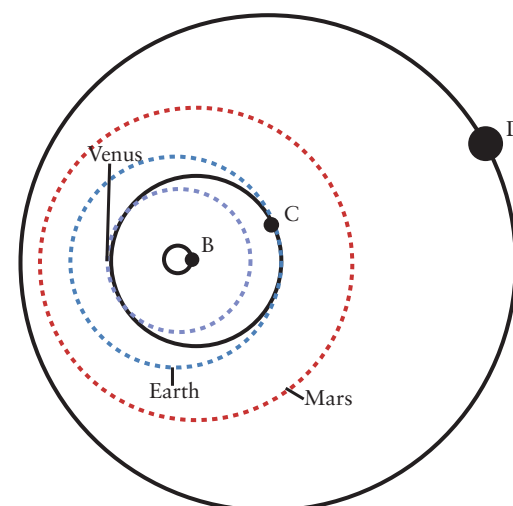
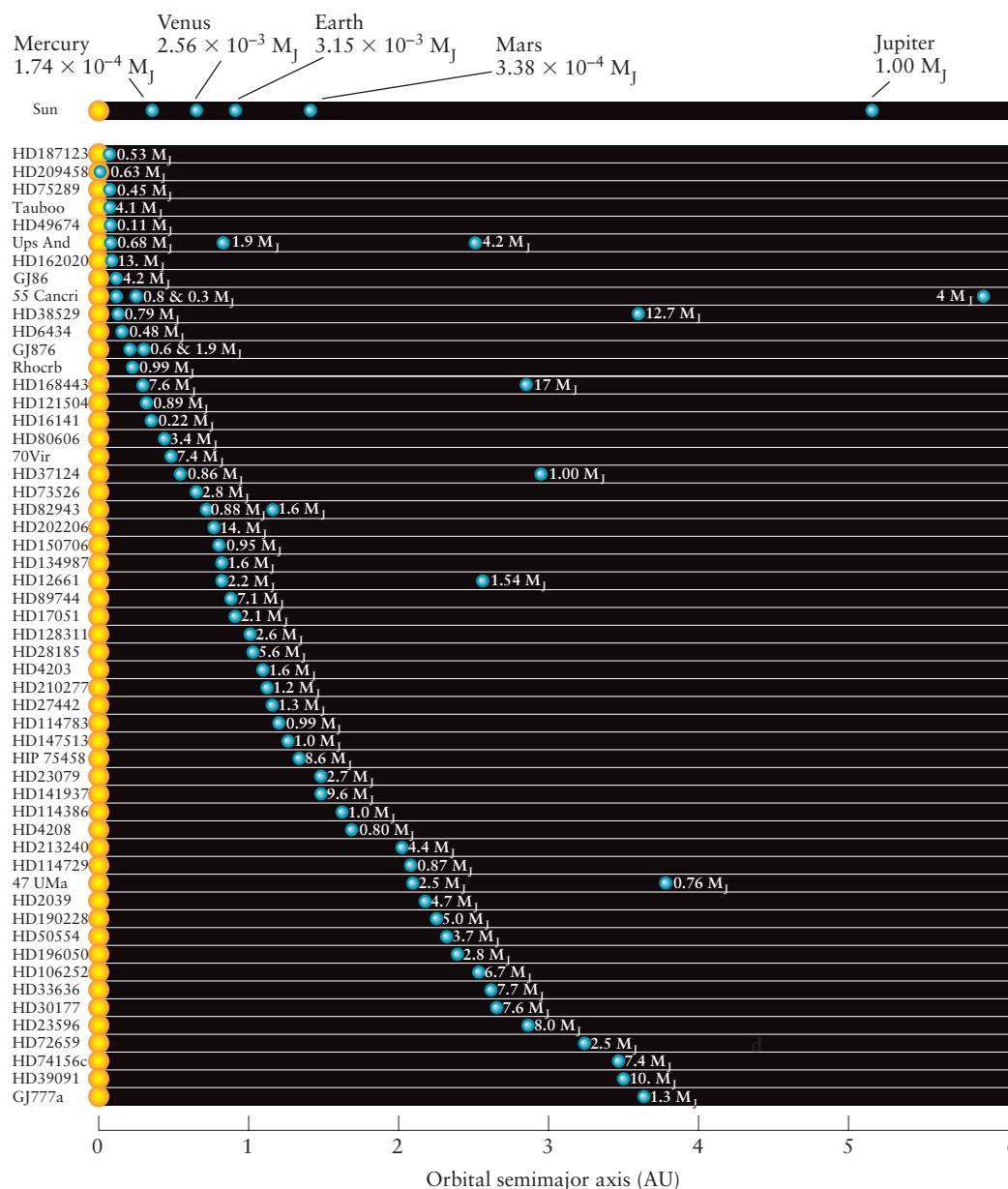
In 2008, a roughly Jupiter-mass planet, WASP-12b, was discovered so close to its star, WASP-12, that tides caused by the star have made the planet egg-shaped. WASP-12b's atmosphere is now being pulled onto its star at a rate of about $15 \times 10^{20} \text{ kg}$ ($2 \times 10^{17} \text{ tons}$) per year. The planet will be devoured within 10 million years.

Conversely, the exoplanet farthest from its star is HIP 77900 b, which orbits its star at a distance of 3200 AU. You can explore many facets of exoplanetology at exoplanet.eu.

5-7 Planets that are not orbiting stars have also been observed

While our discussion so far centers on planets orbiting stars, astronomers have begun observing Jupiter-sized planets that are unconnected to stars. Named, for now, *free-floating planets*, *interstellar planets*, *rogue planets*, *orphan planets*, or *nomad planets*, they are believed to have formed around stars and then were ejected by gravitational effects of other planets or nearby stars.

Although only a few free-floating planets have been observed, their very existence allows astrophysicists to estimate how frequently planets get expelled from orbit around stars. Combined with the result that Earth-mass planets should be ejected even more commonly than higher-mass planets, calculations suggest that there may be more free-floating planets in the Milky Way than there are stars.



R I V U X G

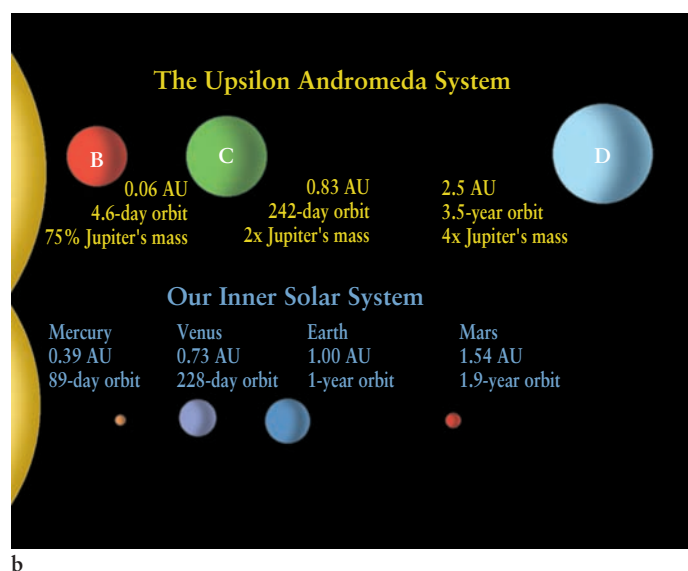


FIGURE 5-8 Planets and Their Stars (a) This figure shows the separations between some exoplanets and their stars. The corresponding star names are given on the left of each line. Note that many systems have giant planets that orbit much closer than 1 AU from their stars. (M_J is shorthand for the mass of Jupiter.) For comparison, the solar system is shown at top. (b) Drawings of the multiple-planet system Upsilon Andromeda. The star Upsilon Andromeda has at least three planets. (c) The orbital paths of the planets, labeled B, C, and D, along with the orbits of Venus, Earth, and Mars, are drawn for comparison. (d) Direct image of four planets orbiting the star HR 8799. (a: Data from California and Carnegie Planet Search. d: National Research Council of Canada; Reprinted by permission from Macmillan Publishers Ltd: *Nature* 468, 1080–1083, copyright 2010)

5-8 There are billions and billions of planets

Given the growing number of exoplanets that are being discovered orbiting a wide range of stars, astronomers are beginning to make estimates for the total number of planets in our Milky Way Galaxy and the number of those planets orbiting Sunlike stars. The results are impressive: On average, every five stars have, among them, eight planets. At least one in every six stars has an Earth-like planet orbiting it closer than Mercury is orbiting the Sun. Our Galaxy is believed to have more planets than stars, both of which run in the hundreds of billions. All of these numbers will be refined as better estimates are determined for the percentage of star systems with planets.

5-9 Planets with liquid water are being discovered

2 We have started finding Earthlike exoplanets orbiting Sunlike stars. Which of these planets will be the best candidates for supporting life as we know it? We are most likely to discover life (of the sort we are familiar

with on Earth) on exoplanets orbiting at about 1 AU from stars similar in mass (within about a factor of 2) to the Sun. At this distance, such stars provide enough heat to keep surface water liquid, and these stars shine long enough for life to evolve. The region around each star where water can be liquid on terrestrial planets in those locations is called the *habitable zone* or *goldilocks zone* (Figure 5-9). The latter name comes about because such regions are not too hot (for water to be liquid) and not too cold. The first Earth-sized planet to be discovered orbiting in its star's habitable zone was Kepler 186f. To date at least 10 Earthlike exoplanets in the habitable zones of Sunlike stars have been discovered.

Not all stars have habitable zones of roughly 1 AU, as we have in our solar system. Stars with different masses have different temperatures and they emit different amounts of heat and light. Therefore, stars with different masses have different habitable zones, as shown in Figure 5-9. Earthlike planets with surface temperatures suitable for liquid water to exist on them must therefore be at different distances from stars with different masses.

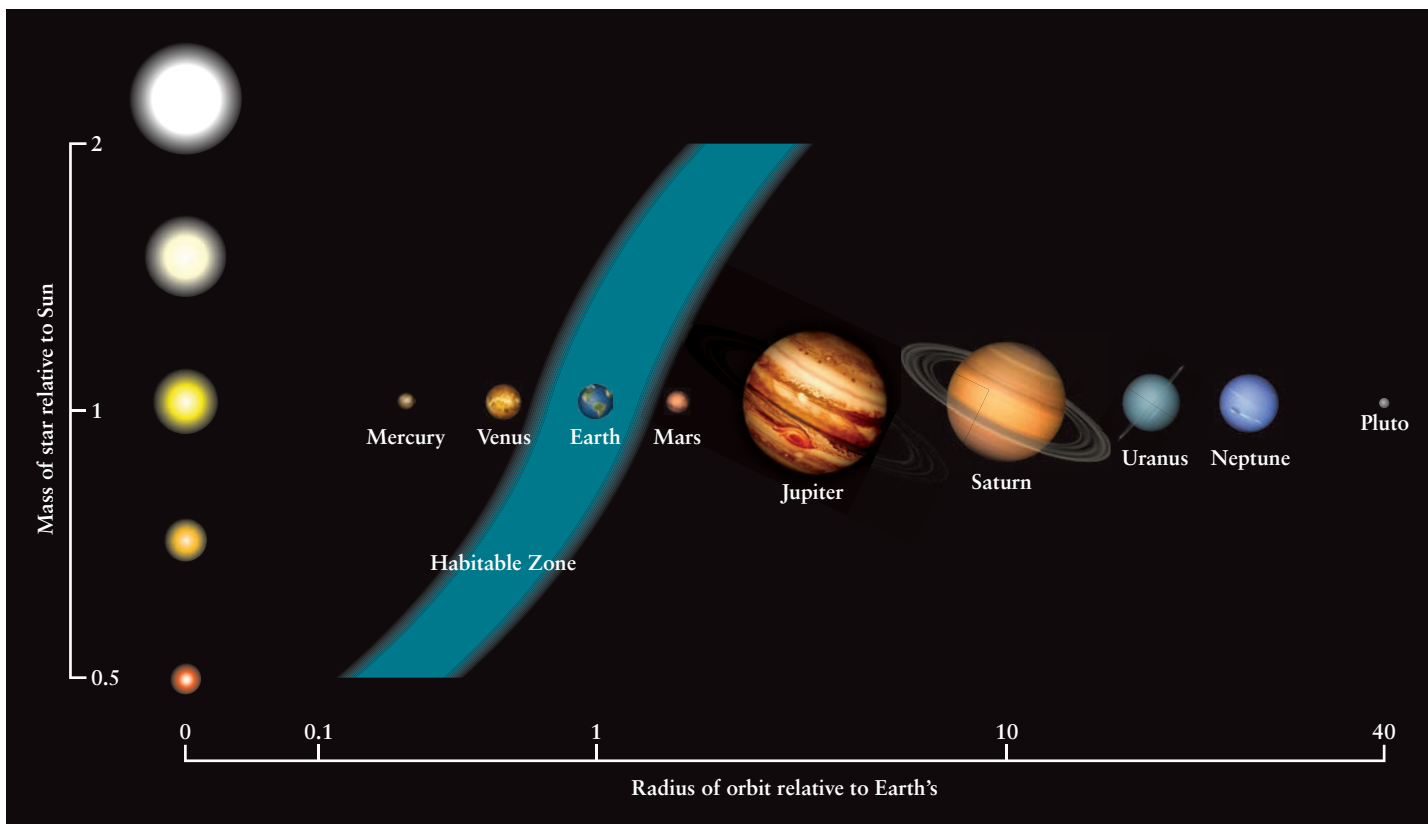


FIGURE 5-9 Habitable Zones around Stars with Different Masses Stars with different masses have different temperatures and they emit different amounts of heat and light. Therefore, Earth-

like planets with surface temperatures suitable for liquid water to exist on them must be at different distances from stars with different masses.

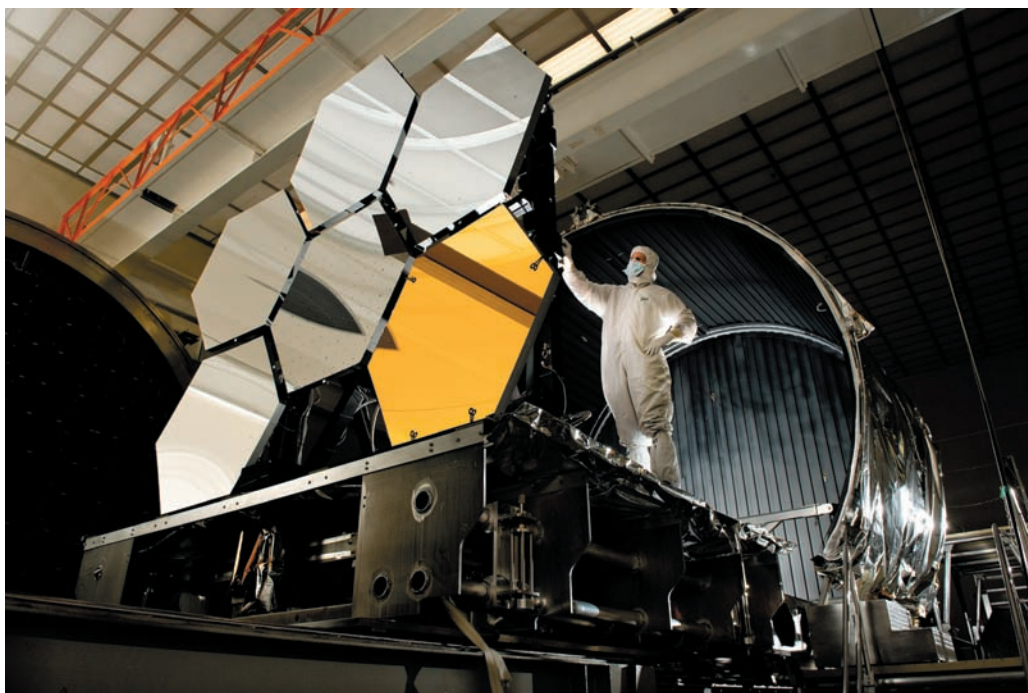


FIGURE 5-10 Aerospace lead optical test engineer Dave Chaney inspects six of the 18 primary mirror segments of NASA's James Webb Space Telescope, prior to cryogenic (extremely low temperature) testing in the X-ray & Cryogenic Facility at NASA's Marshall Space Flight Center in Huntsville,

Alabama. The James Webb Space Telescope will be launched in 2018 to study the formation of the first stars and galaxies and shed new light on the evolution of the universe. (NASA/MSFC/David Higginbotham)

5-10 The search for life on exoplanets is under way

The presence of life on Earth has transformed both the atmosphere and surface of our planet. Early plant life transformed Earth's atmosphere from one that was primarily composed of carbon dioxide and nitrogen to the present atmosphere composed primarily of nitrogen and oxygen. No other mechanism for such a radical transformation is known consistent with the fact that lifeless Venus still has the carbon dioxide and nitrogen atmosphere that it developed 4.5 billion years ago. Life has changed the surface of our planet as a result of forests growing up, cities being built, and land being transformed for agricultural use, among other reasons.

3 To date, no life as we know it has been observed on any exoplanet. If life exists on Earthlike planets orbiting in the habitable zones of other stars, we expect to find similar transformations of the exoplanet atmospheres and surfaces as have occurred here. Therefore, astronomers are beginning to study the atmospheres of exoplanets in the hope of finding oxygen-rich ones. These observations are made as starlight passes through or scatters off exoplanet atmospheres. By studying the spectra of this starlight, we can determine the major components of the planetary atmospheres. These observations will be greatly

accelerated when the James Webb Space Telescope is launched in 2018 (Figure 5-10). That telescope's sensitivity to infrared radiation will far eclipse any other such technology we presently have and will allow us to survey many planetary atmospheres in detail.

The search for advanced technological civilizations on exoplanets began in 1960, long before the first exoplanet was discovered. Cornell astronomer Frank Drake used a 26-m telescope at the National Radio Astronomy Observatory's facility in Green Bank, West Virginia, to listen for radio signals that extraterrestrials may be sending. All of the radio signals he detected were random noise. The search for extraterrestrial-generated radio signals is at the heart of the present Search for Extraterrestrial Intelligence (SETI). To date, no radio signals from any intelligent extraterrestrial source have been detected. The search continues! We have more to say about life on other worlds in Chapter 15.

SUMMARY OF KEY IDEAS

- Astronomers have observed disks of gas and dust orbiting young stars.
- At least 1800 exoplanets have been discovered orbiting other stars.

- Exoplanets ranging in mass from less than the mass of Earth to many times the mass of Jupiter have been detected.
- Most of the exoplanets that have been discovered have masses roughly equal to the mass of Jupiter.
- Exoplanets of sizes ranging from 0.2 to 25 Earth diameters have been observed.
- Some exoplanets are observed directly, while most are detected indirectly as a result of their effects on the stars they orbit.
- Exoplanets orbiting virtually all types of stars have been observed.
- There are more planets than stars in our Milky Way Galaxy.
- Exoplanets supporting life must be Earthlike in composition and in their stars' habitable zones.
- No life has been discovered on any exoplanet.
- Some planets that are not orbiting stars have been observed.

WHAT DID YOU THINK?

- 1 *Have any planets been observed orbiting other stars?* Yes, at least 1800 exoplanets have been observed.
- 2 *Have any Earthlike planets been discovered orbiting in the habitable zones around other stars?* Yes, at least 10 Earthlike exoplanets have been observed in their stars' habitable zones.
- 3 *Have astronomers discovered evidence for life on any planets orbiting other stars?* No.

Review Questions

1. Describe how exoplanets are discovered by measuring stellar radial velocities.
2. Describe how exoplanets are discovered by observing stellar motion across the celestial sphere.
3. Describe how exoplanets are discovered by measuring stellar brightnesses.
4. Describe how exoplanets are discovered by microlensing.
5. Describe two methods astronomers use to search for life on exoplanets.
6. What is a planet's habitable zone?

Advanced Question

7. How would a planet orbiting one of the first-generation stars be different than planets formed today? What chemical elements would that planet contain? Could we humans exist on that world? Why or why not?

Got It?

8. Most extrasolar planets found to date are most similar in mass and size to
 - a. Earth
 - b. Mars
 - c. Jupiter
 - d. Pluto
 - e. our Moon
9. Are all exoplanets orbiting stars? If not, where are they?
10. How are exoplanets found from the Doppler shifts of stars?



R I V U X I G Earth and the Moon as seen from space. (NASA)

CHAPTER

6

The Terrestrial Planets and Their Moons

WHAT DO YOU THINK?

- 1 Can Earth's ozone layer, which has been partially depleted, be naturally replenished?
- 2 Do we see all parts of the Moon's surface at some time throughout the lunar cycle?
- 3 Does the Moon rotate and, if so, how fast?
- 4 What causes the ocean tides?
- 5 Which terrestrial planet—Mercury, Venus, Earth, or Mars—has the coolest surface temperature?
- 6 Which planet is most similar in size to Earth?
- 7 Which terrestrial planet—Mercury, Venus, Earth, or Mars—has the hottest surface temperature?
- 8 Is life known to exist on Mars today?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

In this chapter we first compare basic properties of all eight planets. Then we explore the terrestrial worlds (Mercury, Venus, Earth, and Mars) in detail. The much larger gas giants, Jupiter and Saturn, and the intermediate-sized ice giants, Uranus and Neptune, are examined in the next chapter.

In this chapter you will discover

- that the Earth’s interior and surface are constantly changing
- how Earth’s magnetic field helps protect us
- what made the craters on the Moon
- how the Sun and the Moon cause Earth’s tides
- that water ice has been found on the Moon
- Mercury, a Sun-scorched planet with dormant volcanoes, a heavily cratered surface, and a substantial iron core
- Venus, perpetually shrouded in thick, poisonous clouds and mostly covered by gently rolling hills with few craters
- Mars, a red, dusty planet that once had running water on its surface and may still have liquid water underground

COMPARATIVE PLANETOLOGY

It often helps our understanding of complex topics, like planets, to see the big picture of how the major bodies are related to each other before exploring these objects in detail. To this end, we begin studying the eight planets in our solar system by comparing their physical properties. With this broad perspective in hand, we will then examine each planet and its accompanying moons individually. Earth and the three planets similar to it in chemistry, Mercury, Venus, and Mars, are studied in this chapter, whereas the larger planets, which are rock and metal (Earthlike) planets surrounded by substantial volumes of hydrogen, helium, and water, are explored in Chapter 7.

6-1 Comparisons of the eight planets show distinct similarities and significant differences

Orbit The planets that newly forming the accretion of planetesimals in the young solar system were

Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. As noted in Section 4-9 and Table 6-1, they all have similar orbital properties in that they all orbit the Sun in the same direction and in nearly the same plane (low orbital inclination). Note that, except for Mercury, their low orbital eccentricities mean that their orbits are nearly circular.

Size The planets fall into three size groups (Figure 6-1). The four inner planets form one group; Jupiter and Saturn, the gas giants, form the second group; and Uranus and Neptune, the ice giants, comprise the third, intermediate-size group. Jupiter is the largest planet, with a diameter about 11 times bigger than that of Earth, whereas Mercury, with about a third of Earth’s diameter, is the smallest planet. Indeed, two moons, Ganymede and Titan, are larger than Mercury. Neptune and Uranus are each about 4 times larger in diameter than Earth, the largest of the four inner planets. The diameters of the planets are given in Table 6-2.

Mass A measure of the total amount of matter an object contains, its mass, is another characteristic that distin-

TABLE 6-1 ORBITAL CHARACTERISTICS OF THE PLANETS

	Average distance from the Sun			Sidereal Orbital period (Earth years)	Orbital eccentricity (<i>e</i>)	Orbital inclination
	(AU)	(10 ⁶ km)	10 ⁶ mi			
Mercury	0.39	58	36	0.24	0.206	7.01°
Venus	0.72	108	67	0.62	0.007	3.39°
Earth	1.00	150	93	1.00	0.017	0°
Mars	1.52	228	142	1.88	0.093	1.85°
Jupiter	5.20	778	483	11.86	0.048	1.30°
Saturn	9.54	1427	887	29.46	0.054	2.48°
Uranus	19.19	2871	1784	84.01	0.047	0.77°
Neptune	30.06	4497	2794	164.79	0.009	1.77°

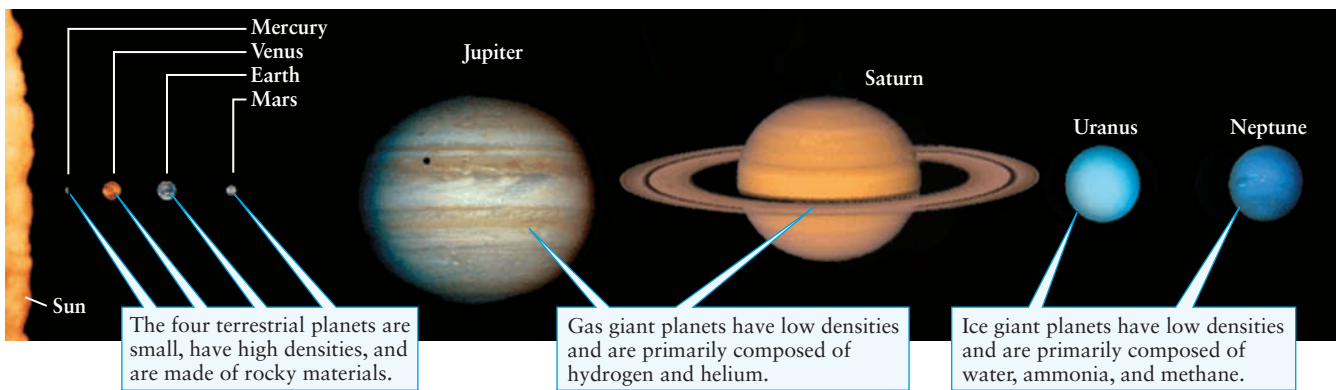


FIGURE 6-1 The Sun and the Planets In this figure, photographs of the eight planets are shown to size scale. From left to right: The four planets that orbit nearest the Sun (Mercury, Venus, Earth, and Mars) are small and made of rock and metal.

The next two planets (Jupiter and Saturn) are large and composed primarily of hydrogen and helium. Uranus and Neptune are intermediate in size and contain roughly equal amounts of ices, hydrogen and helium, and terrestrial material. (NASA)

guishes the inner planets from the outer planets. The unit of mass in the metric system we use in this book is the kilogram. A slight complication occurs in describing mass in U.S. customary units. In those units the standard of mass is the slug, which is rarely used. Rather, we use the other unit of mass in U.S. customary units, the pound (lb, sometimes called the pound-mass). The confusion arises because the pound is also the common unit of weight. Weight is the force with which an object pushes down on a scale, while mass, as just mentioned, is a measure of the number and kind of particles an object possesses. In this text, pound will indicate mass when it is used fol-

lowing units of kilograms (kg) or when used in describing density (lb/yd^3). Planetary masses are determined by measuring the periods of moons' orbits or, for planets without moons, the gravitational effect of the planets on the orbits of passing objects. This information is used in Kepler's third law (or similar equations) to determine the planets' masses. The four inner planets have small masses compared to the giant outer ones. Again, first place goes to Jupiter, whose mass is 318 times greater than Earth's mass (see Table 6-2). Uranus and Neptune, ice giants, have masses intermediate between the terrestrial and the gas giant planets.

TABLE 6-2 PHYSICAL CHARACTERISTICS OF THE PLANETS

	Diameter			Mass			Average density	
	(mi)	(km)	(Earth = 1)	(lb)*	(kg)	(Earth = 1)	(lb/yd ³)	(kg/m ³)
Mercury	3031	4878	0.38	7.3×10^{23}	3.3×10^{23}	0.06	9150	5430
Venus	7519	12,100	0.95	10.8×10^{24}	4.9×10^{24}	0.81	8850	5250
Earth	7926	12,756	1.00	13.2×10^{24}	6.0×10^{24}	1.00	9300	5520
Mars	4216	6786	0.53	14.1×10^{23}	6.4×10^{23}	0.11	6660	3950
Jupiter	88,846	142,984	11.21	4.2×10^{27}	1.9×10^{27}	317.94	2240	1330
Saturn	74,898	120,536	9.45	12.6×10^{26}	5.7×10^{26}	95.18	1160	690
Uranus	31,763	51,118	4.01	19.2×10^{25}	8.7×10^{25}	14.53	2170	1290
Neptune	30,775	49,528	3.88	2.2×10^{26}	1.0×10^{26}	17.14	2760	1640

* Pound (lb), as used here and most other places throughout this book, is a measure of mass, rather than weight.

Focus Question 6-1

Venus and Mercury do not have any moons. What objects do you think astronomers observed being deflected in order to find the mass of these planets?

Density Size and mass can be combined in a useful way to provide information about the chemical composition of a planet (or any other object). Matter composed of heavy elements, such as iron or lead, has more particles (protons and neutrons) and, hence, more mass packed into the same volume than does matter composed of light elements,

such as hydrogen, helium, or carbon. Therefore, objects made primarily of heavier elements have a greater average density than objects composed primarily of lighter elements, where average density is given by the equation

$$\text{Average density} = \frac{\text{Total mass}}{\text{Total volume}}$$

The chemical composition (kinds of elements) of any object determines its **average density**—how much mass the object has in a unit of volume. For example, a kilogram of rock takes up less space (is denser) than a kilogram of sugar, even though both have the same mass. Average density in metric units is expressed in kilograms per cubic meter (kg/m^3). Common practice in the United States is to define density as pound/volume, such as lb/yd^3 , where pound here denotes mass, as discussed in the section on mass. For example, water has a density in metric units of 1000 kg/m^3 , which corresponds to roughly 1700 lb/yd^3 in U.S. customary units. We will always present metric, and usually U.S. customary, numbers throughout the book.

To make the density of astronomical objects more comprehensible, it is often helpful to mentally compare these values with that of water, given above. The four inner planets have high average densities compared to water (see Table 6-2). In particular, the average density of Earth is 5520 kg/m^3 (9300 lb/yd^3). Because the density of typical surface rock is only about 3000 kg/m^3 (5100 lb/yd^3), this high density implies that Earth contains large amounts of material inside it, such as iron and nickel, which are much denser than surface rock.

In sharp contrast, Jupiter, Saturn, Uranus, and Neptune have relatively low average densities (see Table 6-2).

Indeed, Saturn's average density is less than that of water. The low densities of the giant outer planets suggest that they contain significant amounts of the lightest elements, hydrogen and helium, as discussed in Chapter 3. As we saw

in Section 4-3, Jupiter and Saturn share similar compositions, being primarily hydrogen and helium, while Uranus and Neptune both contain vast quantities of water and ammonia, as well as much hydrogen and helium. These four worlds are sometimes called *Jovian planets* (the Roman god Jupiter was also known as Jove), because, like Jupiter, they all have relatively low densities and high masses compared to Earth. However, the differences in the chemistries of Jupiter and Saturn compared

to those of Uranus and Neptune make the name *giant planets* (giant compared to Earth) more appropriate in collectively describing these four. As noted earlier, Jupiter and Saturn are specifically called gas giants, while Uranus and Neptune are ice giants.

Spectra Further information about the chemical composition of bodies in the solar system is obtained from their spectra. For solar system objects, spectra are provided primarily by sunlight scattered off their surfaces or the clouds that surround them. As we saw in Chapter 3, spectra provide us with details of an object's surface (or atmospheric) chemical composition and rotation rate. From their spectra, we can confirm that the outer layers of the giant planets are primarily hydrogen and helium and that the surface of Mars is rich in iron oxides, among many other things.

Albedo The surfaces or the upper cloud layers of the various planets scatter (send in all directions) different amounts of light. The fraction of incoming light returning directly into space is called a body's **albedo**. An object that scatters no light has an albedo of 0.0; for example, powdered charcoal has an albedo of nearly 0.0. An object that scatters all of the light that strikes it (a high-quality mirror comes close) has an albedo of 1.0. Multiplying the albedo of an object by 100 gives the percentage of light directly scattered off that body. Three planets (Mercury, Earth, and Mars) have albedos of 0.37 or less. Such albedos result from dark, dry surfaces or from a mixture of light (water and clouds) and dark surfaces (continents) exposed to space. In contrast, Venus, Jupiter, Saturn, Uranus, and Neptune have albedos of 0.47 or more. High albedos like these imply bright materials exposed to space. The albedos of these latter five planets are high because they are all completely enshrouded by clouds, which scatter sunlight back into space very efficiently.

Moons Every planet, except Mercury and Venus, has moons. At least 173 planetary moons are known to

Focus Question 6-2

Why is Earth's albedo continually changing?

Insight Into Science**Astronomical Measurement**

Astronomers use the laws of physics to infer things we cannot measure directly. For example, we can get an overall idea of the chemical compositions of planets from their masses and volumes: Kepler's laws enable us to determine the mass of a planet from the period of its moons' orbits. The measured diameter of each planet (determined by its distance from Earth and its angular size in the sky) yields its volume. As shown in the equation earlier in this section, dividing total mass by total volume yields the average density. Comparing this density to the densities of known substances gives us information about the planets' interior chemistries.

exist in the solar system (up from 99 known in 2001), and more are still being discovered. Table 6-3 lists the numbers of moons orbiting each planet. Unlike our Moon, most moons are irregularly shaped, and look more like potatoes than spheres. We will discover that there is as much variety among moons as there is among planets.

EARTH: A DYNAMIC, VITAL WORLD

Figure 6-2 summarizes Earth’s important physical and orbital properties. The symbol ⊕ in Figure 6-2 and hereafter is astronomers’ shorthand for Earth. The other planets have different symbols that we will encounter later.

TABLE 6-3 NUMBER OF MOONS KNOWN TO BE ORBITING EACH PLANET

Mercury	0
Venus	0
Earth	1
Mars	2
Jupiter	67
Saturn	62
Uranus	27
Neptune	14

EARTH: Vital Statistics

Planet symbol:	⊕
Average distance from the Sun:	1.000 AU = 1.496×10^8 km = 9.296×10^7 mi
Maximum distance from the Sun:	1.017 AU = 1.521×10^8 km = 9.45×10^7 mi
Minimum distance from the Sun:	0.983 AU = 1.471×10^8 km = 9.14×10^7 mi
Orbital eccentricity:	0.017
Average orbital speed:	29.79 km/s = 18.5 mi/s
Sidereal period of revolution (year):	365.26 days (1.00 year)
Sidereal rotation period:	0.997 day
Sidereal rotation period (day):	1.00 day = 24.0 h
Inclination of equator to orbit:	23.5°
Diameter (equatorial):	12,756 km = 7925 mi
Mass:	5.974×10^{24} kg = 13.2×10^{24} lb (1 M _⊕)
Average density:	5520 kg/m ³ = 9300 lb/yd ³
Albedo (average):	0.31
Escape speed:	11.2 km/s = 7.0 mi/s
Surface temperature range:	Maximum 60°C = 140°F = 333 K
Mean:	14°C = 57°F = 287 K
Minimum:	−90°C = −130°F = 183 K
Atmospheric composition (by number of molecules):	78.08% nitrogen (N ₂), 20.95% oxygen (O ₂), 0.035% carbon dioxide (CO ₂), about 1% water

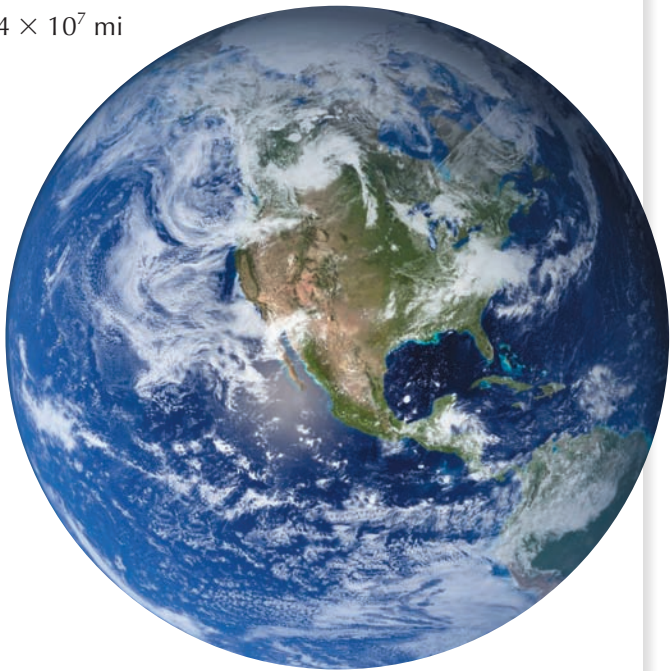


FIGURE 6-2 Earth’s Vital Statistics The planet symbol for Earth is ⊕. Other planets have different symbols; these are all used as shorthand for information about each world. For example, the mass of Earth is denoted M_⊕. This figure provides in-

formation about Earth’s physical and orbital properties. Recall that pounds (lb) here are a unit of mass. Astronauts spend many hours watching our world as they orbit it. The image here gives you an idea of why they find it so fascinating. (NASA)

6-2 Earth's atmosphere has evolved over billions of years

Earth's atmosphere is a key part of our planet's dynamic activity and its ability to sustain life. Essential to the existence of complex life is air that contains oxygen molecules. The air we breathe is predominantly a 4-to-1 mixture of nitrogen to oxygen, two gases that are found only in small amounts in the atmospheres of other planets.

The theory of planet formation and the models that attempt to reproduce the early history of Earth indicate that the present atmosphere is the third one to envelop our planet. Earth's earlier atmospheres contained no free oxygen. Composed of trace remnants of hydrogen and helium left over from the formation of the solar system, the first atmosphere did not last very long. The atoms and molecules in these gases were too light to stay near Earth. Heated by sunlight (meaning that they moved faster and faster), these particles gained enough kinetic energy to overcome Earth's gravitational pull and escape into space.

The gases of the second atmosphere came primarily from inside Earth, vented through volcanoes and cracks in Earth's surface. That atmosphere was composed primarily of carbon dioxide and water vapor, along with some nitrogen and other gases. All these gases consist of more massive atoms than the hydrogen and helium in the first atmosphere. Therefore, the speeds imparted to the gases of the second (and third) atmospheres by sunlight were not high enough to enable them to leave Earth's gravitational bonds in large quantities. That is why Earth developed a permanent atmosphere.

There was roughly 100 times as much gas in that second atmosphere as there is in the air today. Carbon dioxide and water store a great deal of energy in the form of heat, helping to create what is called the *greenhouse effect*, about which we will say more shortly. The high concentration of these gases early in Earth's history stored more heat than the atmosphere does today. Earth back then would have been a serious hothouse had the Sun reached its full intensity at that time. However, the Sun was dimmer and cooler in those days, so the high concentrations of carbon dioxide and water in the air actually prevented the young Earth from freezing.

Oceans began forming when water precipitated out of this atmosphere, as well as from impacts of water-rich space debris, such as the comet nuclei mentioned in Chapter 4. These events occurred within 300 million years of Earth's formation. Indeed, geologists have discovered evidence for liquid water on Earth's surface 4.3 billion years ago. The oceans today cover 71% of Earth's surface. That number can be misleading. Considering that the average ocean depth is only 4.3 km (2.7 mi) and that Earth is 12,756 km (7926 mi) in diameter, the total volume of water on Earth's surface is not that large. One can get a perspective of the total volume of surface water Earth has by comparing the spheres of our planet's solid and liquid components (**Figure 6-3**).

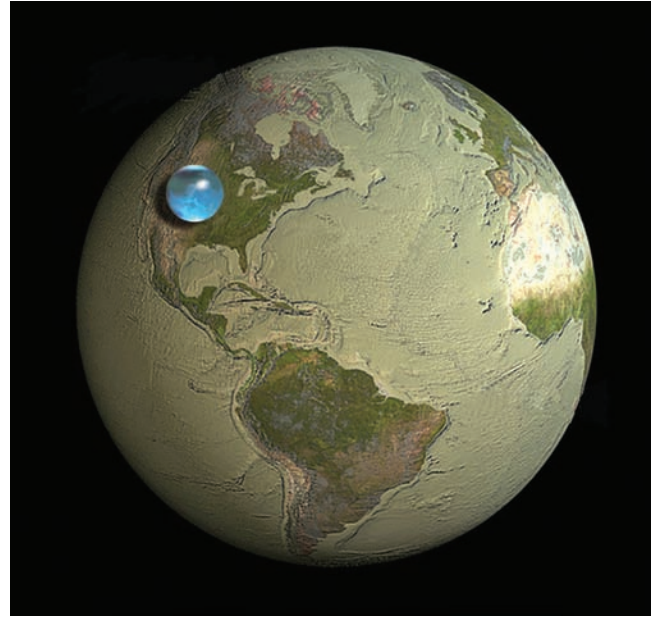


FIGURE 6-3 Earth's Dynamic Oceans This image shows the relative size of Earth compared to a sphere containing all of Earth's water. The sphere is about 1380 km (860 mi) in diameter and includes freshwater, oceans, ice, and even water in the atmosphere. Nearly three-quarters of Earth's surface is covered with water, a substance that is essential to the existence of life. In contrast, there is no liquid water at all on the surfaces of Mercury, Venus, Mars, or the Moon. (Illustration by Jack Cook, Woods Hole Oceanographic Institution)

The oceans eventually soaked up about half the carbon dioxide in that second atmosphere, as the rain absorbed it from the air and carried it earthward. (Water holds a great deal of carbon dioxide, as you can see by the amount of carbon dioxide fizz in a can of soda.) Much of the carbon dioxide in the young oceans combined with water, silicon-based molecules, and calcium-based molecules to create clays, shells, and coral. This material settled to the ocean bottoms, was compressed, and eventually formed rock, like limestone. Therefore, most of the carbon dioxide captured by the oceans has been permanently removed from the atmosphere.

Early plant life in the oceans and then on the land removed most of the remaining carbon dioxide in the air, which plants used to create organic molecules that they needed in order to grow. The most efficient of such conversion mechanisms is photosynthesis, in which oxygen is released into the atmosphere as a waste product. The molecules that remained in the air as the carbon dioxide was being removed were mostly nitrogen, argon, and water vapor. Argon interacts very little with other atoms—it is an inert gas.

Oxygen that entered the air very early in Earth's history came from oxygen-rich molecules, like carbon dioxide, that were broken apart by the Sun's ultraviolet radiation. Then came oxygen released as a by-product of early plant life activity, as just described. In both cases,

that oxygen did not stay in the air for long. Oxygen is highly reactive, and early atmospheric oxygen combined quickly with many elements on Earth's surface, notably iron, to form new compounds, such as iron oxides (otherwise known as rust).

About 2 billion years ago, after all of the surface minerals that could combine with oxygen had done so, the atmosphere began to fill with this gas. With the carbon dioxide reduced to a very low level, the air became the primarily nitrogen-oxygen mixture that we breathe today. Although nitrogen was only a tiny fraction of that earlier atmosphere, it now dominates the thinner air of today.

Because air has mass, it is pulled down toward Earth by gravity, thereby creating a pressure on the surface. Recall that we define pressure as a force acting over some area:

$$\text{Pressure} = \frac{\text{Force}}{\text{Area}}$$

The weight of the air pushing down creates a pressure at sea level of 14.7 pounds per square inch, which is commonly denoted as 1 atmosphere (atm). That value is the pressure acting on your skin. The atmospheric pressure decreases with increasing altitude, falling by roughly half with every gain of 5.5 km (3.4 mi) of altitude. Seventy-five percent of the mass of the atmosphere lies within 11 km (approximately 7 mi, or 36,000 ft) of Earth's surface.

Lower layers of the atmosphere All of Earth's weather—clouds, rain, sleet, and snow—occurs in the lowest layer of the atmosphere, called the **troposphere**. If Earth were the size of a typical classroom globe, the troposphere would be only as thick as a sheet of paper wrapped around its surface.

Starting at Earth's surface, atmospheric temperature initially decreases as you go upward, reaching a minimum of 218 K (−67°F) at the top of the troposphere. Above this level is the region called the **stratosphere**, which extends from 11 to 50 km (approximately 7 to 31 mi) above Earth's surface (Figure 6-4). This region is the realm of the **ozone layer**.

The ozone layer The ozone layer has been much in the news in recent years. What is it, and how threatened is it? Ozone molecules (three oxygen atoms bound together, O₃) are created in the stratosphere when short-wavelength ultraviolet radiation from the Sun strikes and splits normal oxygen molecules, O₂, into two separate oxygen atoms. Each resulting oxygen atom then combines with different O₂ molecules to create ozone, O₃. Once created, ozone efficiently absorbs intermediate-wavelength solar ultraviolet rays, thereby heating the air in this layer while preventing much of this lethal radiation from reaching Earth's surface.

The troposphere (the lowest part of the atmosphere) is heated primarily by Earth, which is why the temperature in this part of the atmosphere decreases with altitude—the farther away from the heat source of Earth, the cooler the air gets. However, because the ozone in the

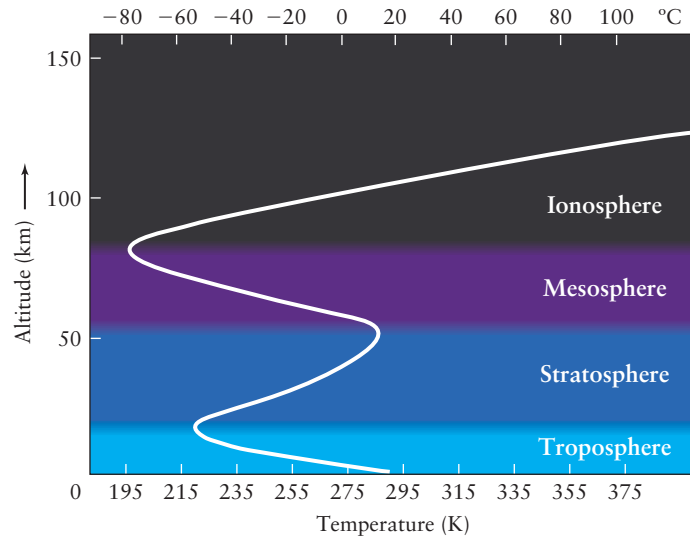


FIGURE 6-4 Temperature Profile of Earth's Atmosphere

The atmospheric temperature changes with altitude because of the way sunlight and heat from Earth's surface interact with various gases at different heights.

stratosphere stores significant heat from the Sun above it, the temperature in the stratosphere actually rises with altitude to about 285 K (50°F) at its top (see Figure 6-4).

Not much ozone actually exists in the ozone layer. If all of the ozone there were compressed to the density of the air we breathe, it would be a layer only a few millimeters (about 1/8 inch) thick. This amount of ozone is enough, however, to protect us from most of the Sun's lethal intermediate-wavelength ultraviolet radiation.

Insight Into Science

What's in a Word? The word “layer” in the term *ozone layer* suggests a narrow region, perhaps a few meters thick. Not so. The ozone layer extends in altitude over some 40 km (25 mi) of the atmosphere!

Ultraviolet radiation has been a double-edged sword in the history of Earth. Because Earth's early atmosphere lacked ozone, more ultraviolet radiation penetrated to the planet's surface, where it provided much of the energy necessary for life to form and evolve in the oceans. Early aquatic life generated O₂, some of which entered the atmosphere. Ultraviolet radiation from the Sun split some of these molecules into separate atoms, thereby enabling the formation of the ozone layer, as described above. As a result, the amount of solar ultraviolet radiation reaching Earth's surface was limited, thereby allowing life-forms to leave the oceans and survive on land.

Today, the stratospheric ozone layer is still vital, because the energy it absorbs would otherwise cause skin cancer and various eye diseases. In sufficiently high doses, ultraviolet radiation even damages the human immune system by destroying some of the molecules that normally repair cells in our bodies. Unfortunately, over

the past few decades, the ozone layer has been partially depleted by human-made chemicals rich in bromine or chlorine, such as chlorofluorocarbons (CFCs). Until the past few years, ozone levels in the mid-latitudes of the northern hemisphere have been decreasing by about 4% per decade. Much more rapid and dramatic drops in the ozone layer over Earth's poles, especially over the South Pole, have been observed for more than two decades. These seasonal *ozone holes* typically extend over 24 million km² (9.3 million mi²), or 4.6%, of Earth's surface area. There are some hints that today the ozone layer is decreasing more slowly and that the ozone holes are diminishing, but these trends are still unclear.

1 Fortunately, as explained earlier, sunlight creates ozone. If current international efforts aimed at banning ozone-depleting compounds (such as some previously common refrigerants) continue, the ozone layer will naturally replenish itself in a century.

There is actually a second layer of ozone that has been developing over the past few centuries, which is hazardous to our health. It is being generated by vehicle and industrial emissions. This ozone is located in the troposphere, along with other air pollution. Unlike the ozone in the stratosphere, the tropospheric ozone is damaging plants and causing respiration problems for animals, including humans. This ozone also contributes to global warming.

Upper layers of the atmosphere Note in Figure 6-4 that above the stratosphere lies the **mesosphere**. Atmospheric temperature again declines as you move up through the mesosphere, reaching a minimum of about 200 K (−103°F) at an altitude of about 80 km (50 mi). This minimum marks the bottom of the **ionosphere**, also called the **thermosphere**, above which the Sun's ultraviolet radiation heats and ionizes atoms, producing charged particles that reflect radio waves.

Carbon dioxide and the greenhouse effect today Perhaps you have had the experience of parking your car in the sunshine on a warm summer day. You roll up the windows, lock the doors, and go on an errand. After a few hours, you return to discover that the interior of your automobile has become considerably hotter than the outside air temperature (Figure 6-5).

What happened to make your car so warm? First, sunlight entered it through the windows. This radiation was absorbed by the dashboard and the upholstery, increasing their temperatures. Because they become warm, your dashboard and upholstery emit infrared radiation, which you detect as heat and which your car windows do not permit to escape. This energy is therefore trapped inside your car and absorbed by the air and interior surfaces. As more sunlight comes through the windows and is trapped, the temperature continues to increase. The same trapping of sunlight warms greenhouses.

The greenhouse effect in our atmosphere works like this. All wavelengths of electromagnetic radiation from the Sun approach Earth's atmosphere. Some of

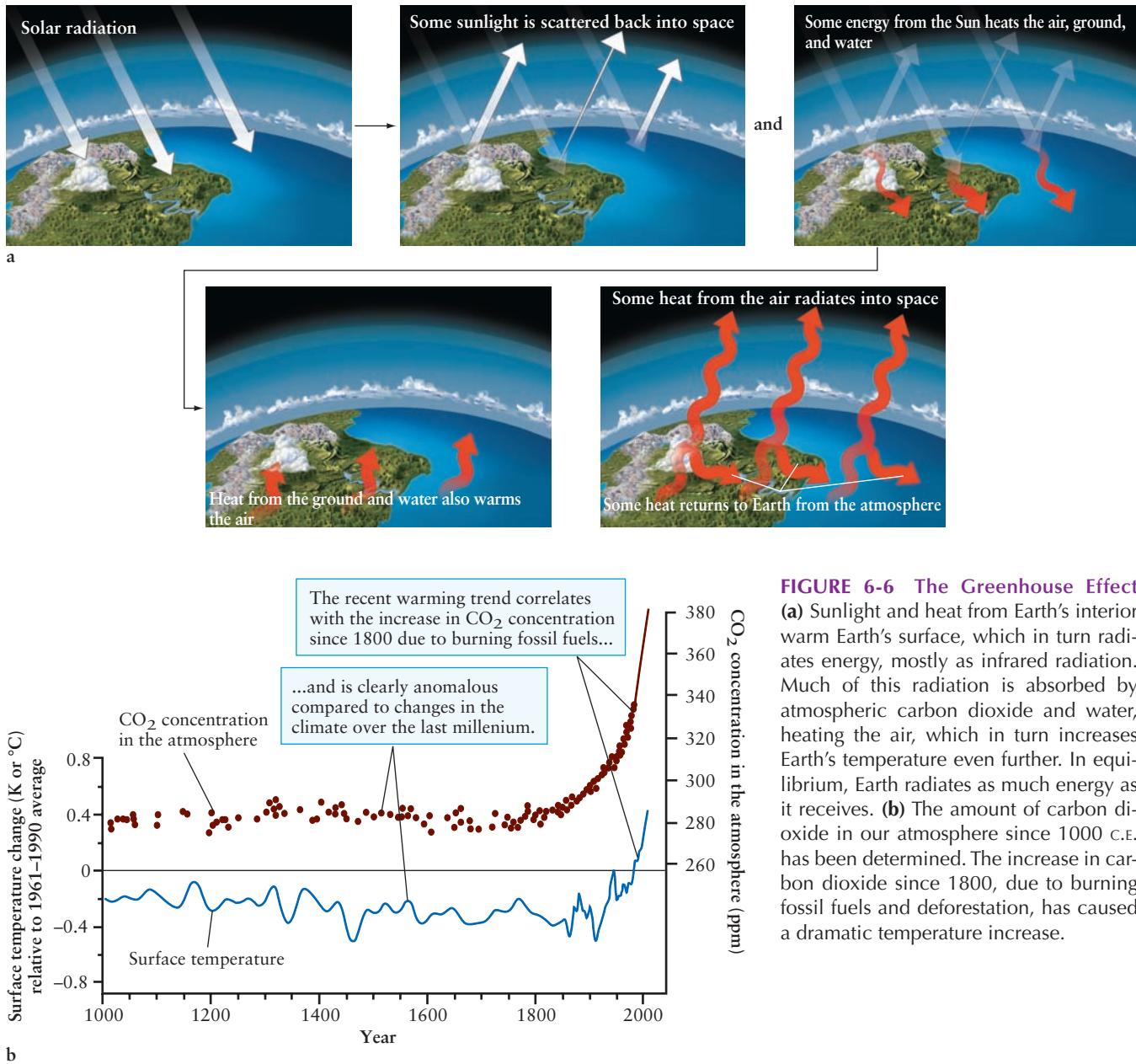


FIGURE 6-5 The Greenhouse Effect in a Car The glass windows in this car allow visible light to enter but prevent the infrared radiation released by the car's interior from escaping. The infrared radiation, therefore, heats the air in the car to a higher temperature than the outside air. This effect is used to advantage in greenhouses and is why it is called the greenhouse effect. (Sheena Goldstein)

this energy is scattered by clouds, land, and water right back into space (Figure 6-6a), while some is absorbed in the atmosphere, thereby heating it. Some visible light, along with some ultraviolet radiation and radio waves, penetrates to Earth's surface and is absorbed by it. This energy helps heat the planet (along with heat from the air and from inside the planet).

Behaving nearly like a blackbody, Earth radiates energy primarily as infrared radiation (where its blackbody spectrum peaks). Unlike most visible light, which passes relatively easily through the gases in our atmosphere, infrared radiation is absorbed by certain gases, especially water, carbon dioxide, and ozone in the air. The energy from these photons further heats the atmosphere. Carbon dioxide levels are increasing in Earth's atmosphere, primarily as a result of human activity, such as the burning of fossil fuels. Increased atmospheric carbon dioxide stores more heat in the air, and the air becomes warmer. Normally, carbon dioxide is removed from the air (and oxygen added) by the process of photosynthesis in green leaves. With increased deforestation and the expansion of cities, fewer plants and trees are left, so the carbon dioxide in the air remains there longer, overheating the air and creating *global warming*. Figure 6-6b shows how the CO₂ concentration has changed over the past millennium.

The consequences of the recent global warming trend are quite complex and serious. They include the melting of glaciers and polar ice caps, rising sea levels, changes in ocean heating and circulation patterns, changes in weather patterns worldwide, and increases in undesirable weather events, such as hurricanes,



tornadoes, droughts, floods, heat waves, and shoreline erosion. With scientific data on the human effects that add to global warming in hand, many nations of the world are working together to decrease its causes.

6-3 Plate tectonics produce major changes on Earth's surface

Large landmasses called continents protrude through Earth's oceans. These landmasses are regions of Earth's outermost layer, or **crust**, that are composed of relatively low-density rock (compared to the rock at the bottom of the oceans). By studying the various kinds of rocks at a particular location on a continent, geologists can deduce

the history of that site. For example, whether an area was once covered by an ancient sea or flooded by lava from volcanoes is readily apparent from the kinds of rocks currently found there.

The crust has distinct boundaries (an example is shown in **Figure 6-7**), many occurring beneath Earth's oceans. These divisions suggest that the continents are separate bodies. Earthquakes and volcanoes often occur near the boundaries, hinting at activity taking place below Earth's crust.

Anyone who carefully examines the shapes of Earth's continents might conclude that landmasses move. Eastern South America, for example, looks like it would fit

Focus Question 6-3

What effect will an increase in the amount of vegetation on Earth have on the carbon dioxide level in the atmosphere?

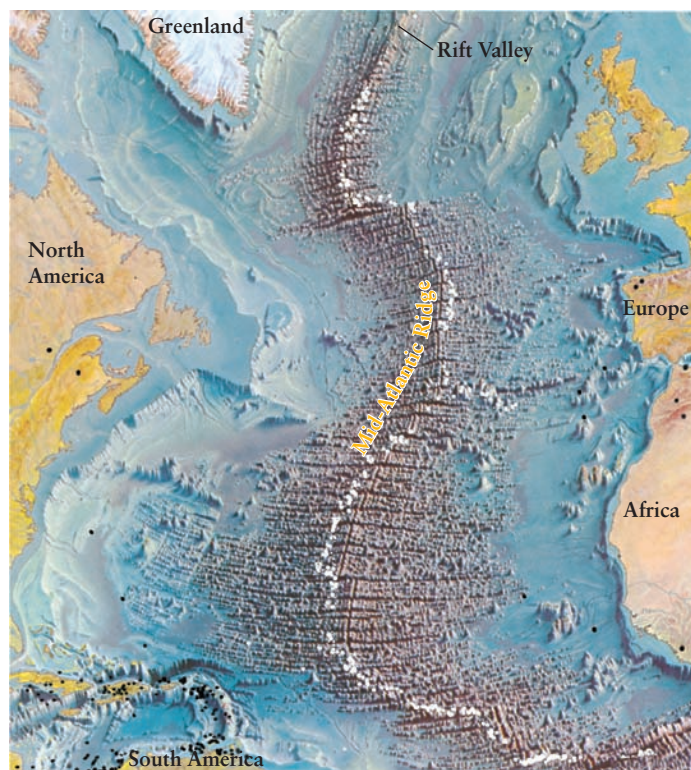


FIGURE 6-7 The Mid-Atlantic Ridge This artist's rendition of the bottom of the North Atlantic Ocean shows an unusual mountain range in the middle of the ocean floor. Called the Mid-Atlantic Ridge, these mountains are created by lava seeping up from Earth's interior along a rift that extends from Iceland to Antarctica. The black and white dots indicate locations of earthquakes. (Buyenlarge/Getty Images)

snugly against western Africa. Between 1912 and 1915, the German scientist Alfred Wegener published his observations on this remarkable fit between landmasses on either side of the Atlantic Ocean, an idea also proposed by Isaac Newton two centuries earlier. Wegener was inspired to advocate the hypothesis of **continental drift**—the idea that the continents on either side of the Atlantic Ocean have moved apart.

Wegener argued that a single, gigantic supercontinent called *Pangaea* (pronounced *pan-gee-yah*, a Greek word for “all lands”) began to break up and drift apart some 200 million years ago. Initially, most geologists ridiculed Wegener's ideas. Although it was generally accepted that the continents do float on denser rock beneath them, few geologists could accept that entire continents move across Earth's face at speeds as great as several centimeters per year. Wegener and other “continental drifters” could not explain what forces would shove the massive continents around.

Then, in the mid-1950s, geologists discovered long, underwater mountain ranges. The Mid-Atlantic Ridge, for example, stretches from Iceland to Antarctica (see Figure 6-6). Careful examination revealed that molten rock from Earth's interior is being forced upward there

Insight Into Science

Acceptance of Scientific Theories Before they are accepted, scientific theories may undergo years or decades of testing and replication. Newton speculated about continental drift in the late seventeenth century. Wegener had proposed continental drift by 1912. His hypothesis was not verified until the 1950s and was not widely accepted until the late 1960s.

as some mechanism causes the ocean floor to slide apart on either side of the ridge. This **seafloor spreading** is moving the Americas away from Europe and Africa at a speed of roughly 3 cm (1.2 in) per year. The relative motion of the plates worldwide ranges from about 1.5 to 7 cm (0.6 to 2.8 in) per year.

Seafloor spreading provided the physical mechanism that allowed for Wegener's hypothesis of crustal motion to be developed into the modern theory of **plate tectonics**. It was after this hypothesis was confirmed in the 1960s that the theory of continental drift became widely accepted.

Earth's surface is made up of about a dozen major thick rafts of rock called *tectonic plates* that move relative to one another. These plates float on the slowly moving, denser matter below them. The plates are not always separate, as they are today. Sometimes they merge into a single mass called a *supercontinent*. In recent years, geologists have uncovered evidence that points to a whole succession of supercontinents that broke apart and reassembled. *Pangaea* is only the most recent supercontinent in this cycle, which repeats on average every 500 million years (Figure 6-8).

The boundaries between tectonic plates are the sites of some of the most impressive geological activity on our planet. Earthquakes and volcanoes tend to occur at these boundaries, where the plates are colliding (at convergent boundaries), separating (at divergent boundaries), or grinding against each other (at *transform-fault boundaries*). If the epicenters of earthquakes are plotted on a map, the boundaries between tectonic plates (Figure 6-9) stand out clearly. For example, the San Andreas fault, running along the western coast of the United States, is the transform fault that occurs where the North American and Pacific plates are rubbing against each other (Figure 6-9b). The Red Sea and Gulf of Suez are expanding as Africa and the Middle East are moving apart (Figure 6-9c). This is an example of a *divergent boundary*. Great mountain ranges are thrust up by collisions between pairs of continent-bearing plates composed of low-density rock. These are called *convergent boundaries*. During such collisions, the colliding boundaries of both plates move up. For example, the Himalayas are being formed today as the Indian-Australian and Eurasian plates collide (Figure 6-9d).



a 237 million years ago: the supercontinent Pangaea



b 152 million years ago: the breakup of Pangaea



c The continents today

FIGURE 6-8 The Supercontinent Pangaea (a) The present continents are pieces of what was once a larger, united body called Pangaea. (b) Geologists believe that Pangaea must have first split into two smaller supercontinents, which they call Laurasia and Gondwana. (c) These bodies later separated into the continents of today. Gondwana split into Africa, South America, Australia, and Antarctica, while Laurasia divided to become North America and Eurasia.

When a continent-bearing plate collides with a plate of higher density, ocean-bottom rock, the denser plate, moves downward (subducts) into Earth under the other (see Figure 6-11). This is currently occurring off the coasts of Japan and Chile.

Countless impacts have occurred in the oceans, which cover most of our planet. Some of the impacting bodies were slowed or obliterated before cratering the ocean floor, but others were large enough or fast enough to hit the floor. Tectonic plate motion helps explain why Earth's oceans have so few craters today. The

resulting craters in the oceans rode on their tectonic plates until they reached a convergent boundary, where they were subducted (drawn underground) and thereby erased.

Impact craters created on the dry landmasses (continents and islands) are exposed to the atmosphere. Wind and water erode them, with silt from rivers helping to cover any remnants. Only the youngest craters on Earth remain visible.

What powers all of this activity? Some enormous energy source must be shifting landmasses, and it must have been doing so for billions of years. The answer to that question leads us to a scientific model of Earth's interior.

6-4 Earth's interior consists of a rocky mantle and an iron-rich core

Geologists have calculated that Earth was composed entirely of liquid rock and metal soon after its formation, about 4.6 billion years ago. The violent impact of space debris, the compression of the interior by the mass above it, and the energy released by the breakup of radioactive elements due to *radioactive decay* and *nuclear fission* heated, melted, and kept the young Earth molten. In radioactive decay, radioactive nuclei emit energy and small particles, such as electrons, protons, neutrons, or helium nuclei. In nuclear fission, the atomic nuclei break into substantial pieces, such as when a uranium atom breaks into a rubidium atom and a cesium atom, while also releasing other particles and energy. Fission also creates the heat used to generate electricity in nuclear power plants and some spacecraft, like the Mars rover *Curiosity*. That process differs from how energy is generated by *nuclear fusion* in the Sun, wherein lighter atoms are fused together, as mentioned in Chapter 4 and discussed in detail in Chapter 9.

Most of the iron and other dense elements sank toward the center of the young, liquid Earth, just as a rock sinks in a pond. At the same time, most of the less dense materials were forced upward toward the surface (Figure 6-10a). This process, called **planetary differentiation**, produced a layered structure within Earth: a very dense central **core** surrounded by a **mantle** of less dense minerals, which, in turn, was surrounded by a thin crust of relatively light minerals (Figure 6-10b).

Differentiation explains why most of the rocks you find on the ground are composed of lower-density elements like silicon and aluminum. Much of the iron, gold, lead, and other denser elements found on Earth today either returned to the surface, emerging through volcanoes and other lava flows, or were deposited by impacts later in the life of Earth. Despite this return of heavier elements to the surface, the average density of crustal rocks is 2700 kg/m^3 (4550 lb/yd^3), which is

Focus Question 6-4

What type of boundary exists between Africa and the Middle East?

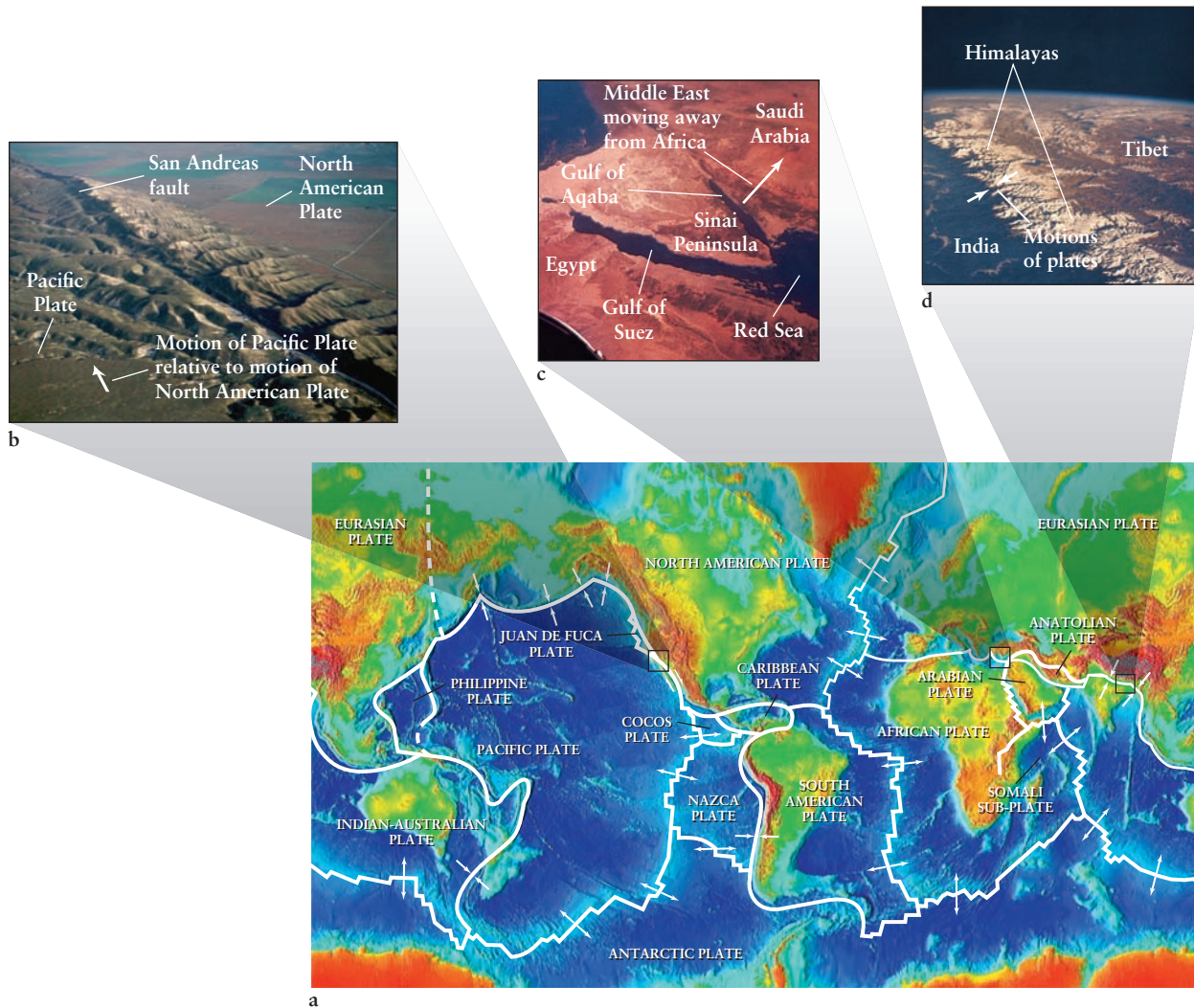


FIGURE 6-9 Earth's Major Tectonic Plates (a) Earth's surface is divided into a dozen or so rigid plates that move relative to one another. The boundaries of the plates are the scenes of violent seismic and geologic activity, such as earthquakes, volcanoes, rising mountain ranges, and sinking seafloors. The arrows indicate whether plates are moving apart ($\leftarrow \rightarrow$), together ($\rightarrow \leftarrow$), or sliding past one another ($\updownarrow$). (b) **Rubbing of Two Plates.** The San Andreas fault, running up the western coast of North America, formed because the Pacific plate is

moving northwest along the North American plate. (c) **Separation of Two Plates.** The plates that carry Egypt and Saudi Arabia are moving apart, leaving the rift valley that contains the Red Sea. (d) **Collision of Two Plates.** The plates that carry India and China are colliding. As a result, the Himalayas are being thrust upward. In this photograph, taken by astronauts in 1968, Mount Everest is one of the snow-covered peaks near the center. (a: Digital image by Peter W. Sloss, NOAA/NESDIS/NGDC; b: Craig Aurness/Corbis; c: *Gemini 12*, NASA; d: *Apollo 7*, NASA)

considerably less than the 5520 kg/m^3 (9300 lb/yd^3) average density of Earth as a whole.

To understand Earth's interior, geologists have created a mathematical model that includes the effects of differentiation, gravity, radioactivity, chemistry, heat transfer, rotation, and other factors. The model predicts that Earth's temperature increases steadily from about 290 K (62°F) on the surface to nearly 6000 K ($10,800^\circ\text{F}$) at the center. Compression of the core, as noted earlier, provides much of the heat that remains there, while radioactivity replenishes most of the heat lost from the mantle through the planet's surface. In addition to temperature, pressure also increases with increasing depth

below Earth's surface because the deeper you go, the more mass presses down on you. These factors shape Earth's inner layers.

You might think that a temperature of 6000 K ($10,800^\circ\text{F}$) would melt virtually any substance. The mantle, in particular, is composed largely of minerals rich in iron and magnesium, both of which have melting points of only slightly more than 1250 K (1790°F) at Earth's surface. However, the melting point of any substance also depends on the pressure to which it is subjected: the higher the pressure, the higher a material's melting point. The high pressures within Earth (1.4 million atm at the bottom of the mantle) keep the mantle relatively solid to

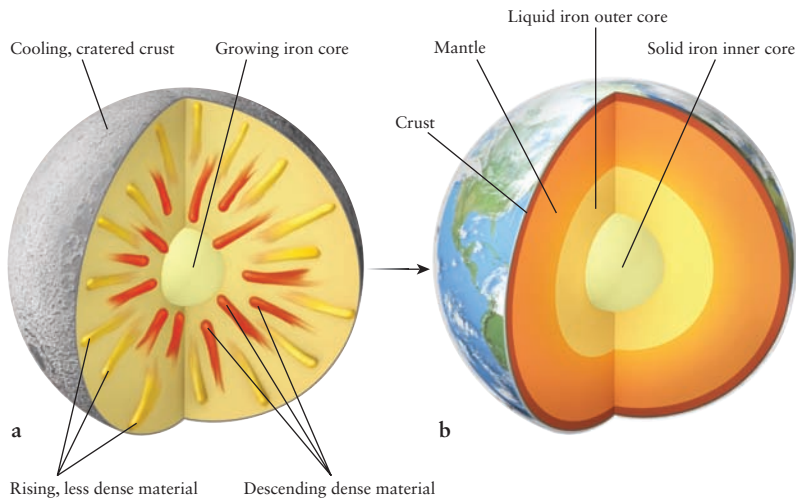


FIGURE 6-10 Cutaway Model of Earth Early Earth was initially a homogeneous mixture of elements with no continents or oceans. **(a)** Molten iron sank to the center and light material floated upward to form a crust. **(b)** As a result, Earth has a dense iron core and a crust of light rock, with a mantle of intermediate density between them.

a depth of about 2900 km (1800 mi). This region still deforms, but more slowly than the rock directly below it, which has the consistency of Silly Putty. The mantle rests on a liquid iron outer core, below which the pressure is so high that it compresses the very hot iron atoms there to form the solid inner core (Figure 6-10b).

Geologists have tested their models of Earth's interior by studying the response of the planet to earthquakes. These events produce a variety of **seismic waves**, vibrations that travel through Earth either as ripples, like ocean waves, or by compressing matter, like sound waves. Geologists use sensitive **seismographs** to detect and record these vibratory motions. The varying density and composition of Earth's interior affect the direction and speed of seismic waves traveling through Earth. By studying the deflection of these waves, geologists have been able to determine the structure of the planet's interior.

Analysis of seismic recordings led to the discovery that Earth's solid core has a radius of about 1250 km (775 mi), surrounded by a molten iron core with a thickness of about 2250 km (1375 mi). For comparison, the overall radius of our planet is 6378 km (3963 mi). The cores are composed of roughly 80% iron and 20% lighter elements, such as nickel, oxygen, and sulfur. The interior of our planet therefore has a curious structure: a liquid region sandwiched between a solid inner core and a solid mantle (see Figure 6-10b).

A sophisticated computer model of Earth's interior predicts that the solid inner core is rotating faster than the rest of the planet. Seismic evidence supports the model that the core spins around about 2° more per

year than the surface. Such observations as well as refined models are revealing more of the secrets of Earth's interior.

The tremendous heat generated inside Earth provides the energy that drives the motion of the tectonic plates in Earth's crust. This heat is transported upward by **convection**. In this process a liquid or gas is heated from below and, as a result, it expands, becomes buoyant, and therefore rises, carrying the heat it received with it. Eventually, the liquid or gas releases the heat, cools, condenses, and sinks back down (subducts). This motion of fluid causes a circular current, called a *convection current*, to arise. You witness convection every time you see liquid simmering in a pot.

Heated from below, blobs of hot fluid or gas move upward. The rising material releases its heat at the surface (Figure 6-11a).

Earth's upper mantle, upon which the tectonic plates lie, is solid rock. However, this rock is weak and hot enough to have an oozing, plastic flow (like stretching Silly Putty). The crust and this upper mantle are collectively called the *lithosphere*. As sketched in Figure 6-11c, molten rock, or *magma*, seeps upward along cracks or rifts in the ocean floor where convection currents of mantle rock are carrying plates apart. Crustal rock sinks back down into Earth where plates collide. The continents ride on top of the plates as they are pushed around by the convection currents circulating beneath them. This convection cycles the rock down and back to the surface every 50 to 200 million years.

Supercontinents like Pangaea apparently sow the seeds of their own destruction by blocking the flow of heat from Earth's interior. As soon as a supercontinent forms, temperatures beneath it increase. As heat accumulates, the supercontinent domes upward and cracks. Overheated molten rock wells up to fill the resulting rifts, which continue to widen as pieces of the fragmenting supercontinent are pushed apart.

Earth's molten, iron-rich interior affects the planet's exterior in another important way, one that leaves fewer obvious signs than tectonic plate motion. This effect is the creation of a magnetic field that extends through Earth's surface and out into space. As we will see next, that magnetic field results from the convection of molten metal in Earth's core combined with the planet's rotation.

6-5 Earth's magnetic field shields us from the solar wind

If you have ever used a compass, you have seen the effect of Earth's magnetic field. This planetary field is quite similar to the field that surrounds a bar magnet

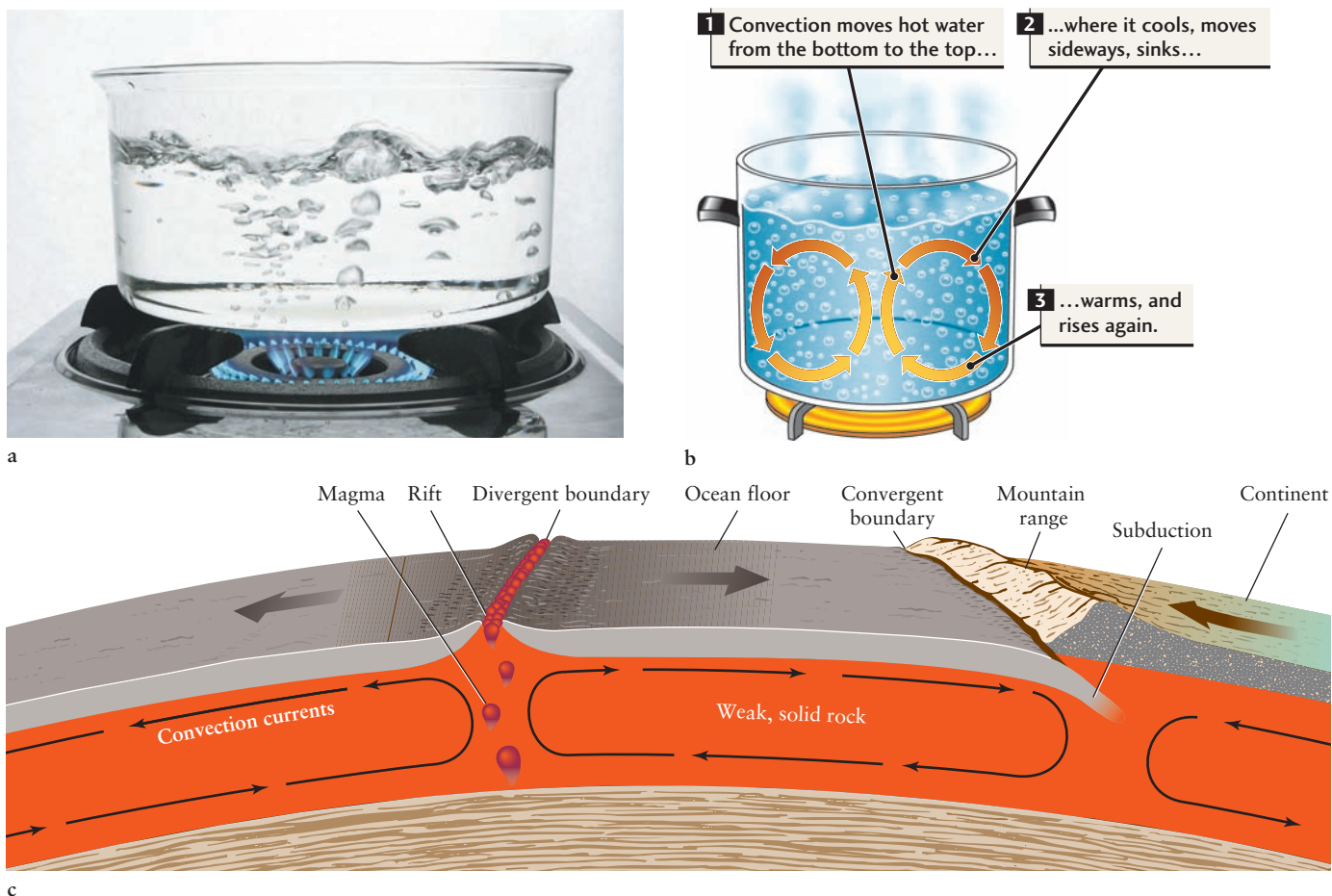


FIGURE 6-11 Convection and the Mechanism of Plate Tectonics (a and b) Heat supplied by burning gas or a heating coil warms the water at the bottom of a pot. The heated water consequently expands, decreasing its density. This lower-density water rises (like bubbles in soda) and transfers its heat to the cooler surroundings. When the hot rising water gets to the top of the pot, it loses a lot of heat into the room, becomes denser, and sinks back to the bottom of the pot to re-

peat the process. (c) Convection currents in Earth's interior are responsible for pushing around rigid plates on its crust. New crust forms in oceanic rifts, where magma oozes upward between separating plates. Mountain ranges and deep oceanic trenches are formed where plates collide, and crust sometimes sinks back into the interior. Note that not all tectonic plates move together or apart—some scrape against each other. (a: Steven Coling/Shutterstock)

(Figure 6-12a). Earth's magnetic field is created by electric charges in motion inside the planet. Analogously, the electric wires in your house also carry moving charges called electric currents that, in turn, create magnetic fields of their own.

Many geologists believe that convection of molten iron in Earth's outer core, combined with our planet's rotation, creates the electric currents, which, in turn, create Earth's magnetic field. The details of this so-called **dynamo theory** of Earth's magnetic field are still being developed.

The magnetic fields near Earth's south rotation pole loop out tens of thousands of kilometers into space and then return near Earth's north rotation pole. The places where the magnetic fields pierce Earth's surface most intensely are called the north and south magnetic poles. These poles move continuously relative to our planet's

rotation poles. Evidence in solidified lava reveals that Earth's magnetic field completely reverses or flips (the north magnetic pole becomes the south magnetic pole and vice versa) on an irregular schedule, typically occurring every 200,000 to 300,000 years. The last reversal is an exception, having taken place about 780,000 years ago. Each flip takes about 4000 years to complete. Today, the magnetic and rotation poles of Earth are about 11.3° apart; some planets have much larger angles between their magnetic and rotation poles. **Figure 6-13** is a scale drawing of the magnetic fields around Earth, which comprise Earth's magnetosphere. Magnetic fields always form complete loops. (The unconnected lines in Figure 6-12b actually are completed outside the boundaries of the figure.)

Our planet's magnetic field protects Earth's surface from bombardment by particles flowing from the Sun

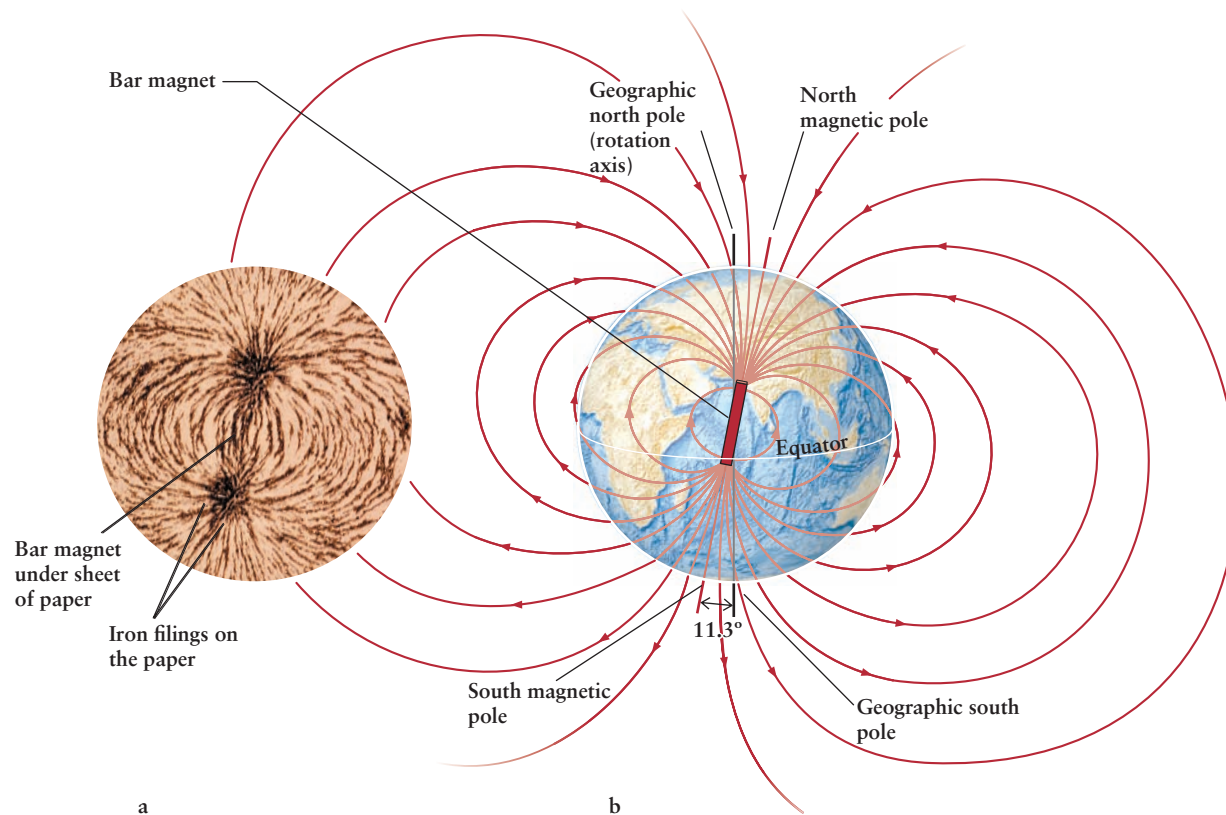


FIGURE 6-12 Earth's Magnetic Field (a) The magnetic field of a bar magnet is revealed by the alignment of iron filings on paper. (b) Although Earth does not contain a bar magnet, its rotation, combined with moving electric charges in its core, creates an equivalent field. Note that the field is not aligned with Earth's rotation axis. By convention, the magnetic pole near

Earth's north rotation axis is called the *north magnetic pole* even though it is actually the south pole on a magnet! We will see similar misalignments and flipped magnetic fields when we study other planets. Note that this figure does not include the distortion of the field by the solar wind, which is depicted in Figure 6-13. (a: Jules Bucher/Photo Researchers/Getty Images)

(see Chapter 9). The *solar wind* is the erratic stream of electrically charged particles ejected from the Sun's upper atmosphere. Near Earth, the particles in the solar wind move at speeds of approximately 400 km/s, or about a million miles per hour.

Magnetic fields can change the direction of moving charged particles. Thus, solar wind particles heading toward Earth are deflected by our planet's magnetic field. Many particles, therefore, pass around Earth (see Figure 6-13). Without Earth's magnetic fields to

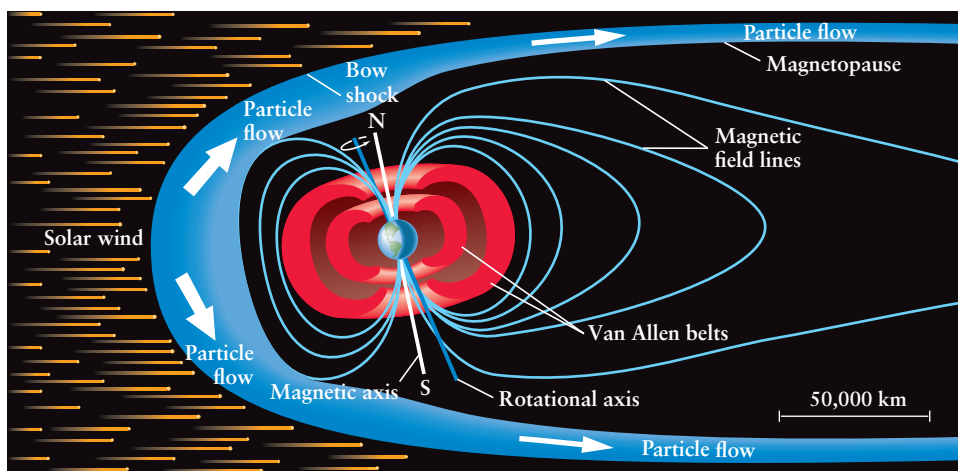


FIGURE 6-13 Earth's Magnetosphere A slice through Earth's magnetic field, which surrounds the entire planet, carves out a cavity in space that excludes most of the charged particles ejected from the Sun, called the solar wind. Most of the particles of the solar wind are deflected around Earth by the fields in a turbulent region colored blue in this drawing. Because of the strength of Earth's magnetic field, our planet traps some charged particles in two huge, doughnut-shaped rings called the Van Allen belts (in red).

deflect them, the solar wind particles would strike air particles, giving many of them enough energy to fly into space and never return. Over billions of years, this would have had the effect of significantly decreasing the thickness of Earth's atmosphere and hence its ability to support life.

Earth's magnetic field also traps some of the solar wind particles in two huge, doughnut-shaped rings called the **Van Allen radiation belts** (hereafter *Van Allen belts*), which surround Earth (see Figure 6-13). The region between the two belts fills up with charged particles, which are drained by lightning on Earth, making this region unstable. These belts, discovered in 1958 during the flight of the first successful U.S. Earth-orbiting satellite, were named after the physicist James Van Allen, who insisted that the satellite carry a Geiger counter to detect charged particles.

When the flow of particles from the Sun is particularly high, some of the solar wind that is trapped in the Van Allen belts is funneled toward Earth's polar regions where these particles descend and cascade down into Earth's upper atmosphere, usually in a ring-shaped pattern (Figure 6-14a). As these high-speed, charged particles collide with gases in the upper atmosphere, the gases fluoresce (give off light) like the gases in a neon sign. The result is a beautiful, shimmering display, called the **northern lights** (*aurora borealis*) or the **southern lights** (*aurora australis*), depending on the hemisphere in which the phenomenon is observed (Figure 6-14b, c).

Focus Question 6-5

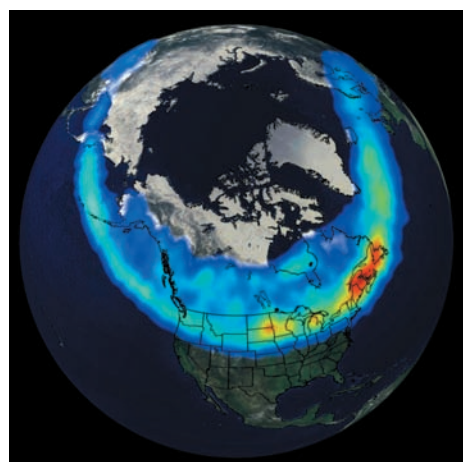
What happens to the Van Allen belts when Earth's magnetic field is flipping?

Insight Into Science

Pace of Discovery Discoveries such as the Van Allen belts remind us how much remains to be learned, even about our own planet. Such fundamental discoveries continue to be made for many reasons. For example, the technology necessary to make certain observations may just have been developed, or our equations that describe various phenomena may become more complete. Sometimes these equations are so complex that their implications are not immediately clear. Progress in science requires unquenchable curiosity and unfailing open-mindedness.

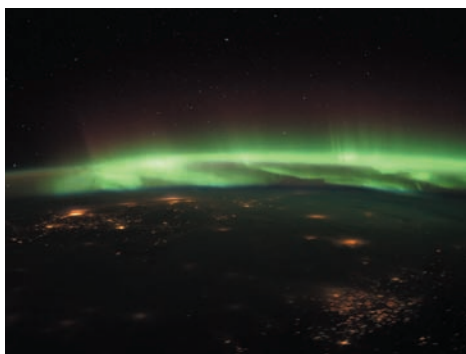
Occasionally, violent events on the Sun's surface, called solar flares and *coronal mass ejections*, send bursts of protons and electrons straight through the Van Allen belts and into the atmosphere. These solar events partially deplete the ozone layer and cause spectacular auroral displays that can often be seen over a wide range of latitudes and longitudes. These solar events also disturb radio, TV, and cell phone transmissions and can damage communications satellites and electric transmission lines on Earth.

In this era of burgeoning satellite usage and high technology here on Earth, the effects of the Sun's electromagnetic radiation, along with the flow of particles from it, are increasingly important. As a result, *space weather* satellites are now continually monitoring our



a

R I V U X G



b

R I V U X G



c

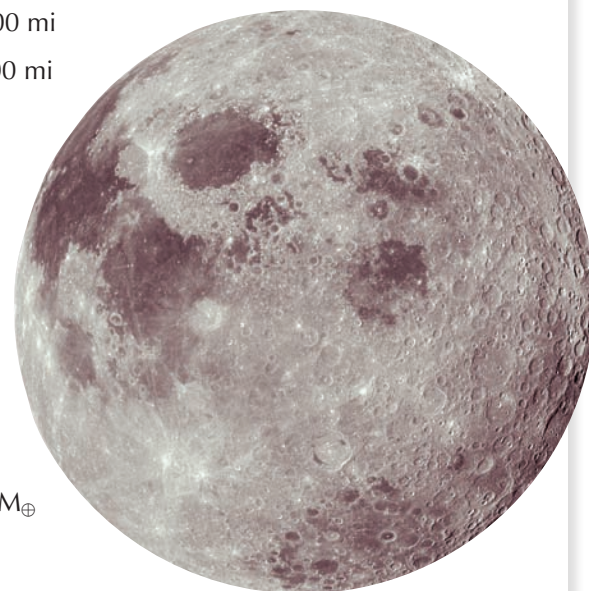
R I V U X G

FIGURE 6-14 The Northern Lights (Aurora Borealis) A deluge of charged particles from the Sun can overload the Van Allen belts and cascade toward Earth, producing auroras that can be seen over a wide range of latitudes. (a) View of an aurora from the *Polar* spacecraft imposed on a graphic image of Earth. Colors from blue to red indicate lowest to highest auroral activity. Auroras typically occur 100 to 400 km (62 to 250 mi) above Earth's surface. (b)

View of aurora over the north central part of the United States and southern Canada as seen from the International Space Station when it was over south central Nebraska. (c) Aurora borealis in Alaska. The gorgeous aurora seen here is mostly glowing green due to emission by oxygen atoms in our atmosphere. Some auroras remain stationary for hours, while others shimmer, like curtains blowing in the wind. (a: Greg Shirah/GSFC/NASA; b: NASA; c: J. Finch/Science Source)

THE MOON: Vital Statistics

Distance from Earth (center to center):	Average: 384,400 km = 238,900 mi Maximum (apogee): 405,500 km = 252,000 mi Minimum (perigee): 363,300 km = 225,700 mi
Orbital eccentricity:	0.055
Synodic period of revolution (lunar phases):	29.53 days
Sidereal rotation period (around Earth):	27.32 days
Inclination of equator to orbit:	6.68°
Inclination of orbit to ecliptic:	5.15°
Diameter (equatorial):	3476 km = 2160 mi = 0.272 Earth diameter
Mass:	7.35×10^{22} kg = 16.2×10^{22} lb = 0.012 $M_{\oplus}$
Average density:	3340 kg/m ³ = 5630 lb/yd ³
Albedo (average):	0.12
Escape speed:	2.4 km/s = 1.5 mi/s
Average surface temperatures:	Day: 130°C = 266°F = 403 K Night: -180°C = -292°F = 93 K
Atmosphere:	Traces of argon (Ar), helium (He), sodium (Na), and potassium (K)
Surface gravity (Earth = 1):	0.17



R I V I U X I G

FIGURE 6-15 The Moon's Vital Statistics Our Moon is one of seven large satellites in the solar system. The Moon's diameter of 3476 km (2158 mi) is slightly less than the distance from

New York to San Francisco. This image shows roughly half the side facing Earth (left) and half the side we never see from here, known as the far side. (NASA)

neighborhood in space and warning about solar events that could imperil humans and machines in space and equipment, like power grids, here on Earth.

THE MOON AND TIDES

Earth's only natural satellite provides one of the most dramatic sights in the nighttime sky. The Moon is so large and so close to us that some of its surface features are readily visible to the naked eye. Even without a telescope, you can easily see dark gray and light gray areas that cover vast expanses of the Moon (**Figure 6-15**).

6-6 The Moon's surface is covered with craters, plains, and mountains

People have dreamed of visiting our nearest neighbor in space for millennia. One of the first movies ever made (1902) was *A Trip to the Moon*, based on Jules Verne's

book *From the Earth to the Moon*. The wonder of actually sending people there for the first time was one of the few things that has ever caused humanity to stop its quarreling long enough to participate, even vicariously, in a great adventure. To date, 12 people have walked on the Moon. Our exploration of it begins with a study of its surface.

Surface Perhaps the most familiar and characteristic features on the Moon are its craters (**Figure 6-16a**). As discussed in Chapter 4 on the formation of the solar system, all lunar craters that we have examined are the result of bombardment by meteoritic material (rather than volcanic activity). Nearly all of these craters are circular, indicating that they were not merely gouged out by moderate-speed rocks, which would create lopsided craters if they struck the surface at an angle. Instead, the high-speed (upward of 180,000 km/h or 112,000 mi/h) collisions with the Moon's surface vaporized the rapidly

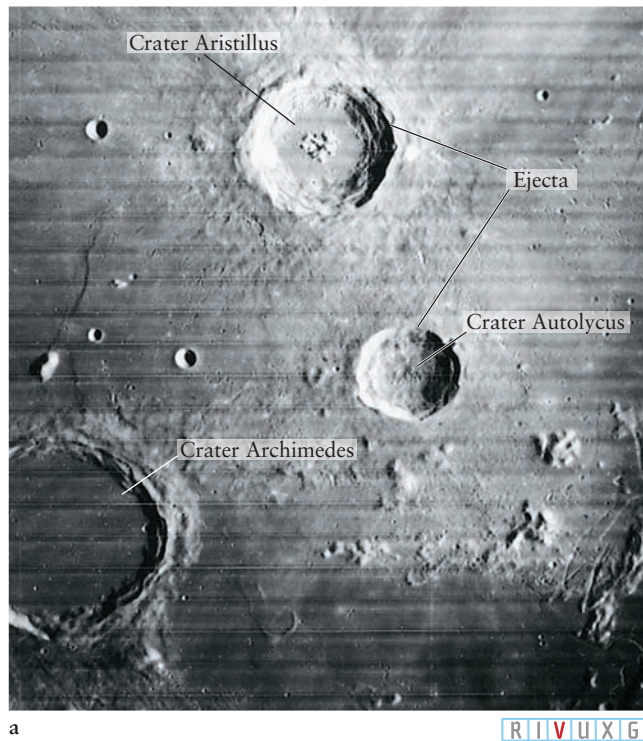
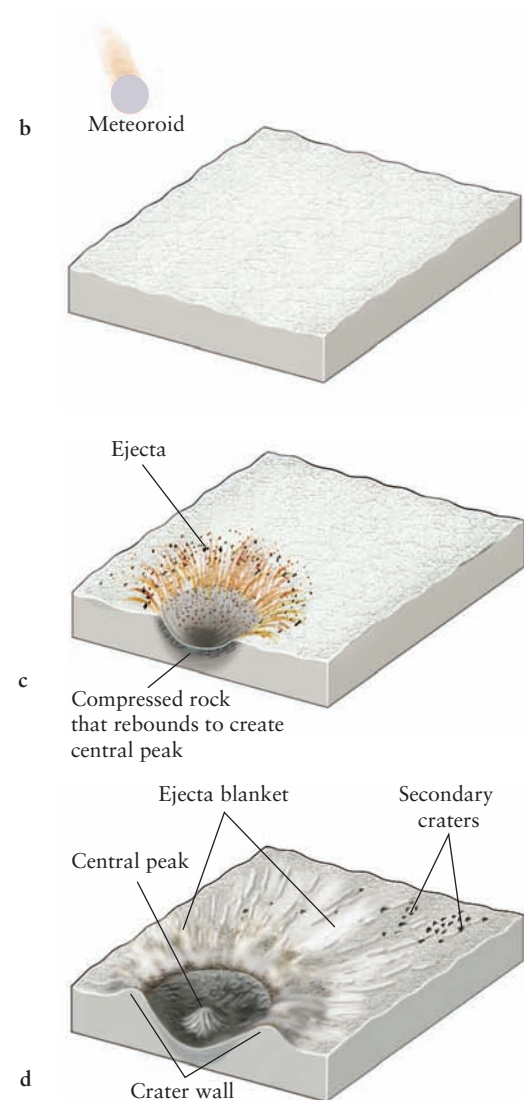


FIGURE 6-16 Lunar Craters (a) This photograph, taken from lunar orbit by astronauts, includes the crater Aristillus. Note the crater's central peaks, the collapsed, terraced crater wall, and the ejecta blanket. Numerous smaller craters resulting from the same impact pockmark the surrounding lunar surface. The three drawings to the right show the crater formation process: (b) An incoming meteoroid, (c) upon impact, is pulverized and the surface explodes outward and downward. (d) After the impact, the ground rebounds, creating the central peak and causing the crater walls to collapse. The lighter region is the ejecta blanket. (a: © 2004 Lunar and Planetary Institute/Universities Space Research Association)



moving debris coming in at any angle, often with the power of a large nuclear bomb. The resulting explosions of hot gas and rubble produced the round-rimmed craters that we observe today (Figures 6-16b, c, and d).

Meteoritic impact causes material from the crater site to be ejected onto the surrounding surface. This pulverized rock is called an **ejecta blanket**. You can see some of the more recent ejecta blankets, which are lighter-colored than the older ones, in Figures 6-15 and 6-16a. If the impact is relatively slow and at a shallow angle (as contrasted with head-on), then the ejecta blanket can have a missing region (Figure 6-17). Laboratory

experiments reveal that the impacting body comes from the direction without ejecta.

Ejecta blankets darken with time as their surfaces roughen from impacts by the solar wind and other particles

from space. Therefore, the lightness of a crater's ejecta is one clue astronomers use to determine how long ago a crater formed.

Craters larger than about 20 km (12 mi) in diameter often form *central peaks* (see Figure 6-16d). These features occur because the impact compresses the crater floor so much that afterward the crater rebounds and pushes the center upward. As the central peak goes up, the crater walls collapse and form *terraces* inside the crater.

Through an Earth-based telescope, some 30,000 craters, with diameters ranging from 1 km (0.62 mi) to more than 100 km (62 mi), are visible. Following a tradition established in the seventeenth century, the most prominent craters are named after astronomers, physicists, mathematicians, and philosophers, such as Kepler, Copernicus, Pythagoras, Plato, and Aristotle. Close-up photographs from lunar orbit reveal millions of craters

Focus Question 6-6

Which of the three named craters in Figure 6-16a is the youngest?

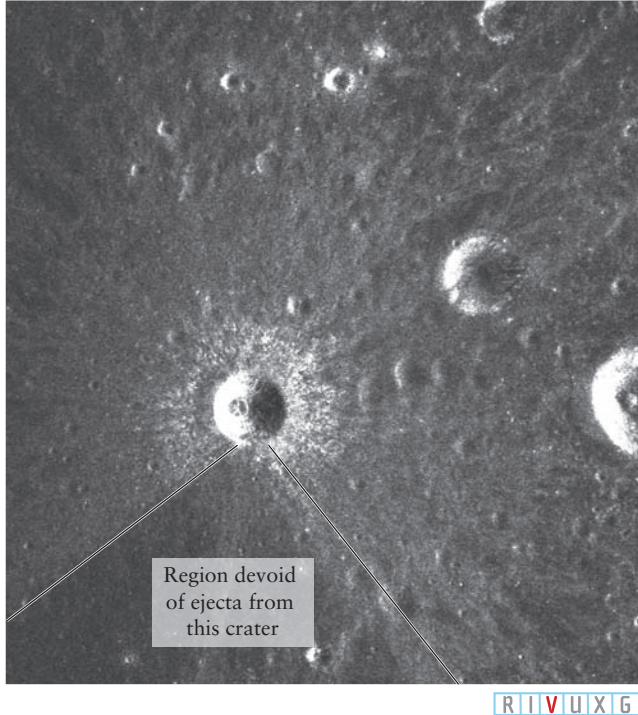


FIGURE 6-17 Shallow Angle Impact Crater This crater, in Mare Nubium, is the result of a very low-angle impact. Despite the missing ejecta between the two lines, the impact crater is still circular, indicating how powerful the impact was. The impacting body came from the direction of the missing ejecta. (NASA)

too small to be seen with Earth-based telescopes. Indeed, extreme close-up photos of the Moon's surface reveal countless microscopic craters (**Figure 6-18**).

Earth, by comparison, has only 178 known craters caused by impacts. Many craters on Earth have been pulled inside our planet by plate tectonic motion and obliterated. Many other craters of a few kilometers in diameter and smaller have been worn away (eroded) by weathering effects of wind and water. Earth's atmosphere vaporized countless space rocks that would otherwise have created small craters here. Moreover,

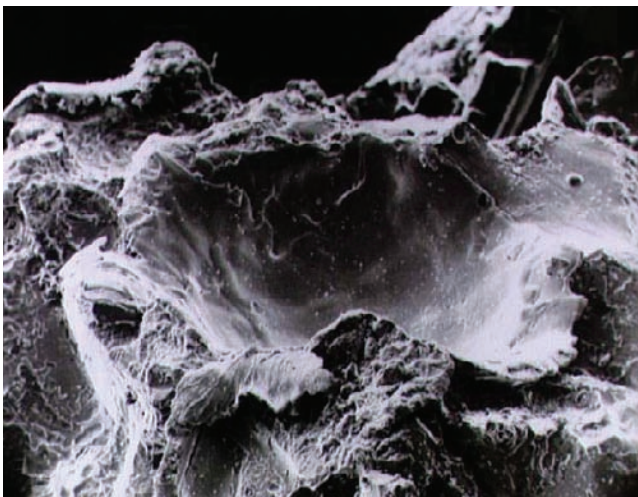


FIGURE 6-18 A Microscopic Lunar Crater This photograph, taken through a microscope, shows a tiny microcrater less than 1 mm (4×10^{-2} in) across on a piece of Moon rock. (NASA)

those that do hit are slowed down by the atmosphere so that they produce smaller craters than they otherwise would have.

Besides its craters, the most obvious characteristic of the Moon visible from Earth is that its surface is various shades of gray (see Figure 6-15). Most prominent are the large, dark gray plains called **maria** (pronounced *mar'-ee-uh*). The singular form of this term, **mare** (pronounced *mar-ay*), means “sea” in Latin and was introduced in the seventeenth century when observers using early telescopes thought that these features were large bodies of water. We know now that no liquid water exists on our satellite. Nevertheless, we retain these poetic, fanciful names, including Mare Tranquillitatis (Sea of Tranquility), Mare Nubium (Sea of Clouds), Mare Nectaris (Sea of Nectar), and Mare Serenitatis (Sea of Serenity). As introduced in Chapter 4 and as we will discuss further shortly, maria are basins on the Moon that were filled in with lava after the Late Heavy Bombardment (see Section 4-7). One of the largest of the maria is Mare Imbrium (Sea of Showers). This mare is roughly circular and measures 1100 km (700 mi) in diameter (**Figure 6-19**).

Although the maria seem quite smooth in telescopic views from Earth, close-up photographs from lunar orbit and from the surface reveal that they contain small craters and occasional wandering channels, called **rilles**

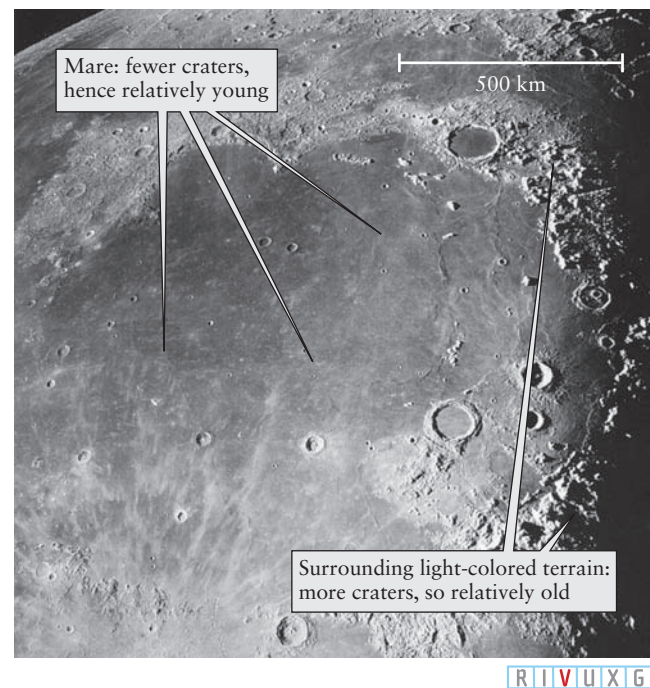
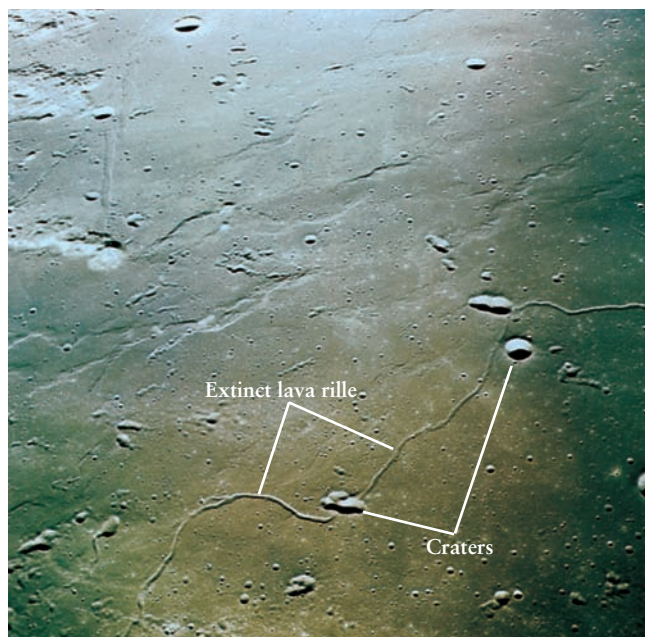


FIGURE 6-19 Mare Imbrium and the Surrounding Highlands Mare Imbrium, the largest of the 14 dark plains that dominate the Earth-facing side of the Moon, is ringed by lighter-colored highlands strewn with craters and towering mountains. The highlands were created by asteroid impacts pushing land together. (NASA)

(Figure 6-20). Rilles are believed to form for at least two reasons: When meandering underground tubes of molten rock stop flowing and the magma in them either drains out or solidifies in place and thereby condenses,

the ceilings of the tubes are no longer supported, and they collapse. Also, if lava rivers flowed on the Moon's surface and then stopped running, they would have drained away, leaving beds that we see as rilles.



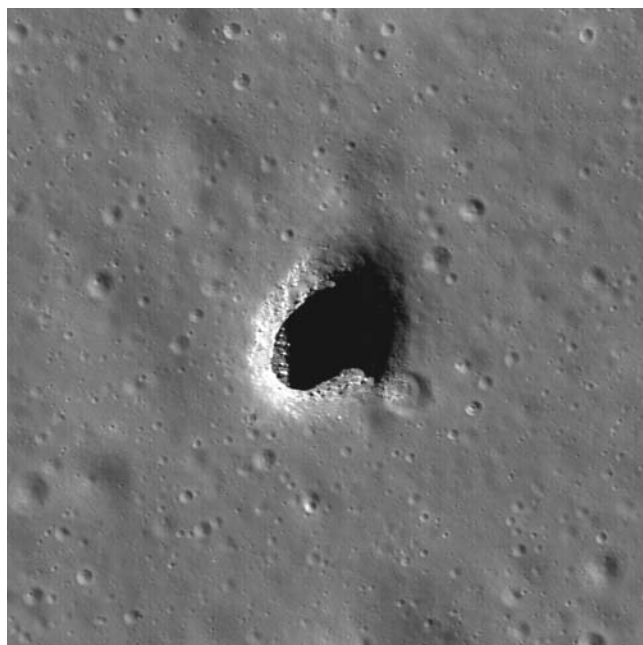
a

R I V U X G



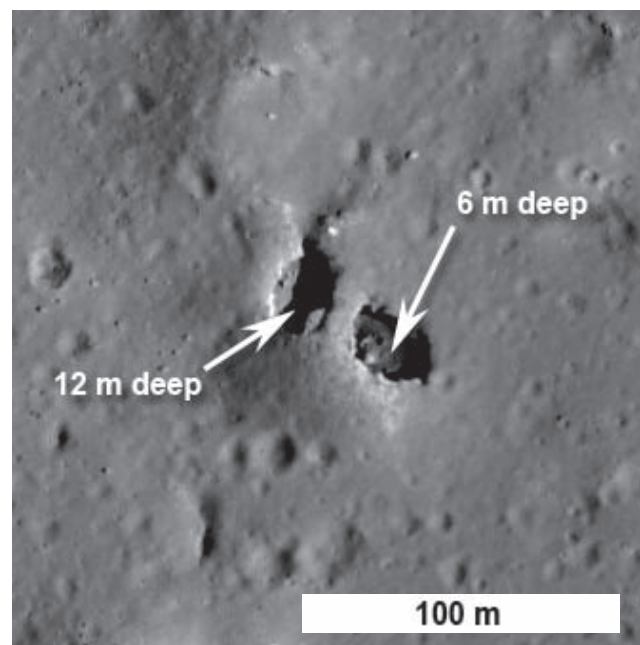
b

R I V U X G



c

R I V U X G



d

R I V U X G

FIGURE 6-20 Details of a Lunar Mare (a) Close-up views of the lunar surface reveal rilles and numerous small craters on the maria. (a) Astronauts in lunar orbit took this photograph of Mare Tranquillitatis (Sea of Tranquility) in 1969 while searching for potential landing sites for the first human landing. At 1100 km (700 mi) across, this mare is the same size as the distance from London to Rome or from Chicago to Philadelphia. (b) Astronaut David Scott on

Hadley's rille during the *Apollo 15* mission to the Moon. (c) Possible lava tube outlet on the Moon imaged by NASA's *Lunar Reconnaissance Observer*. (d) Bridge near King crater. We know this is a bridge, rather than two depressions, because sunlight, coming from the lower right, is visible in the upper left depression. (a: NASA; b: Kennedy Space Center/NASA; c: NASA/Goddard/Arizona State University; d: NASA/GSFC/Arizona State University)

Support for the existence of lava tubes on the young Moon came in 2009, when a Japanese lunar orbiter, *SELENE*, observed a nearly circular hole about 65 m (71 yd) in diameter and 85 m (93 yd) deep on the Moon's surface. This width-to-height ratio is inconsistent with it being an impact crater. Rather, it is believed to be the opening of one such tube (Figure 6-20c). A similar feature, discovered in 2010, is a natural bridge between two parts of the same collapsed opening. The surface apparently solidified before the liquid rock underneath drained away. As the gap widened below the solid surface, some but not all of it collapsed, leaving the bridge (Figure 6-20d). It is interesting to note that emptied lava tubes are also found on Earth in Hawaii, for example.

As we will discuss further in Section 6-8, the same side of the Moon always faces Earth. One of the surprises stemming from the earliest lunar exploration is that the far side is much more heavily cratered than the near side. A few tiny maria, including Mare Moscoviense, Mare Ingenii, and Mare Australe, plus one moderate-sized mare, Orientale, grace the Moon's far side, the side we never see from Earth (Figure 6-21). Unlike the maria on our side, those on the far side are only partially filled with lava. Otherwise, craters and mountains cover the entire far side of the Moon.

Detailed measurements by astronauts in lunar orbit demonstrated that the maria on the Moon's Earth-facing

side are 2 to 5 km (1 to 3 mi) *below* the average lunar elevation on that side, whereas the ones on the far side are 4 to 5 km (2 to 3 mi) *above* the average lunar elevation. The difference in heights of the maria on the two sides of the Moon apparently stems from the way they were formed. Those maria on the near side were created by flows of lava that filled low-lying basins (enormous craters) surrounded by **highlands**. These mountains were pushed up by the large impacts that formed these craters during the Late Heavy Bombardment. The small maria on the far side were more likely lava flows that pushed up and out into craters relatively high in the mountains, rather than into low-lying basins.

The relative ease with which maria on the near side of the Moon formed, compared to the far side, is explained by the observation that the crust on the near side is thinner than on the far side. The impacts that created what became the maria formed craters that thinned the near side crust even further, making eventual seepage of lava easier on our side.

The flat, low-lying, dark gray maria cover only 17% of the entire lunar surface. The remaining 83% is composed of light gray, heavily cratered, mountainous highlands (see Figure 6-15). One mountain range created by an impact can be seen around the edge of Mare Imbrium (see Figure 6-19). Although they take their names from famous terrestrial ranges, such as the Alps, the Apennines, and the Carpathians, the lunar mountains were not formed through plate tectonics, as were mountains on Earth. You can visit images of many regions on the Moon, including the *Apollo* landing sites, at <http://lroc.sese.asu.edu/images>. Many of the images there have the same detail or better than Google maps show of Earth.

6-7 Visits to the Moon yielded invaluable information about its history

Reaching the Moon became national obsessions for the United States and the former Soviet Union in the early 1960s. A series of successful space missions by the Soviet Union spurred President John F. Kennedy's goal of landing a person on the Moon by the end of that decade.

The first of six manned lunar landings, *Apollo 11*, set down on Mare Tranquillitatis on July 20, 1969. Astronaut Neil Armstrong was the first human to set foot on the Moon. In the same year, *Apollo 12* also landed in a mare. Human voyages into space encountered a setback when *Apollo 13* experienced a nearly fatal explosion en route to the Moon. Fortunately, a heroic effort by the astronauts and ground personnel brought it safely home. *Apollo 14* through *Apollo 17* took on progressively more challenging lunar terrain. The period of human visitation to our Moon lasted less than 4 years, culminating (and ending) in 1972, when

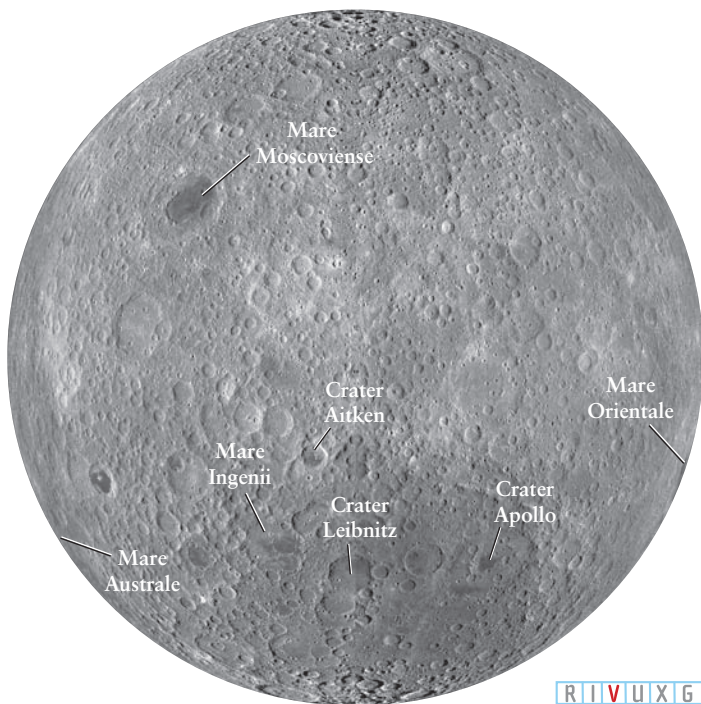


FIGURE 6-21 The Far Side of the Moon Unlike the side of the Moon facing Earth, the lunar far side has only a few small maria. It is almost entirely covered with highlands. (NASA/GSFC/Arizona State University)



R I V U X G

FIGURE 6-22 An Apollo Astronaut on the Moon *Apollo 17* astronaut Harrison Schmitt enters the Taurus-Littow Valley on the Moon. The enormous boulder seen here slid down a mountain to the right of this image, fracturing on the way. This final Apollo mission landed in the most rugged terrain of any Apollo flight. (NASA)

Apollo 17 landed amid mountains just east of Mare Serenitatis (**Figure 6-22**).

Astronauts discovered that the lunar surface is covered with fine powder (**Figure 6-23a**) and rock fragments (**Figure 6-23b**). The powdered rock, ranging in thickness from 1 to 20 m (1 to 22 yd), is called the



a

R I V U X G

FIGURE 6-23 The Regolith The Moon's surface is covered by a layer of (a) powdered rock and (b) small pieces of rock. Called *regolith*, the powdered rock was created over billions

of years as a result of bombardment by space debris; it sticks together like wet sand, as illustrated by this *Apollo 11* astronaut footprint. (a: NASA; b: Apollo 11/NASA)

regolith (from the Greek, meaning “blanket of stone”). Regolith formed during 4.5 billion years of relentless micrometeorite bombardment that pulverized the surface rock. We do not refer to this layer as soil, because the term *soil* suggests the presence of decayed biological matter, which is not found on the Moon's surface.

A major goal of the *Apollo* program was to study the geology of the Moon and help determine its history. The astronauts brought back a total of 382 kg (842 lb) of lunar rocks, which have provided important information about the early history of the Moon. By carefully measuring trace amounts of radioactive elements, geologists determined that lunar rock from the maria is solidified lava dating back only 3.1 to 3.8 billion years. This indicates that the maria developed about a billion years after the Moon formed and that the maria were formed from molten rock, consistent with this material having poured out of the Moon after the Late Heavy Bombardment. These Moon rocks are composed mostly of the same minerals that are found in volcanic rocks in Hawaii and Iceland. The rock of these low-lying lunar plains is called **mare basalt** (**Figure 6-24**). About 17% of the Moon's surface rock is basalt.

In contrast to the dark mare basalt, the lunar highlands are covered with **anorthosite**, rock that is somewhat lighter than basalt (**Figure 6-25**). On Earth, anorthositic rock is found only in very old mountain ranges, such as the Adirondacks in the eastern United States. Compared to the mare basalts, which have more of the heavier elements like iron, manganese, and titanium, anorthosite is rich in calcium and aluminum. Anorthosite is, therefore, less dense than basalt (ignoring the air bubbles in the mare basalts as seen in **Figure 6-24**). Typical



b

R I V U X G

of years as a result of bombardment by space debris; it sticks together like wet sand, as illustrated by this *Apollo 11* astronaut footprint. (a: NASA; b: Apollo 11/NASA)

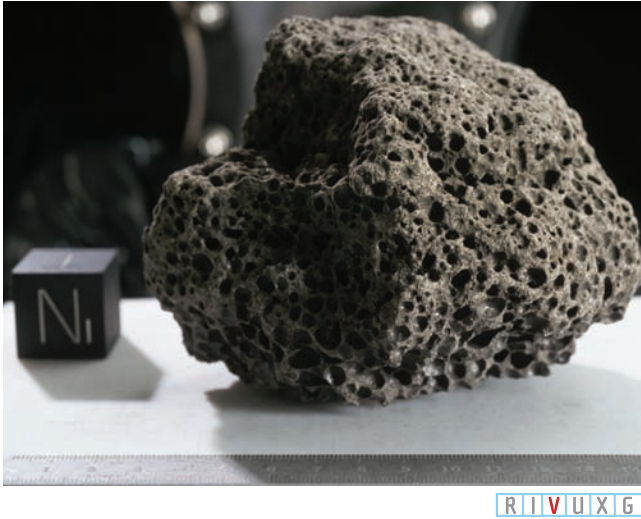


FIGURE 6-24 Mare Basalt This 1.53-kg (3.38-lb) specimen of mare basalt was brought back by *Apollo 15* astronauts in 1971. Small holes that cover about a third of its surface suggest that gas was dissolved in the lava from which this rock solidified. When the lava reached the airless lunar surface, bubbles formed as the pressure dropped and the gas expanded. Some of the bubbles were frozen in place as the rock cooled. (NASA)

anorthositic specimens from the highlands are between 4.0 and 4.3 billion years old. These ancient specimens of highland rock probably represent samples of the Moon's original (pre-Late Heavy Bombardment) crust. Many highland rocks brought back to Earth are **impact breccias**

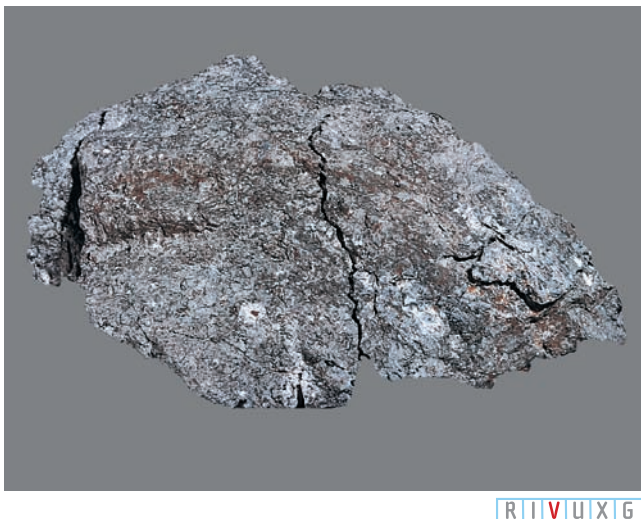


FIGURE 6-25 Anorthosite The lunar highlands are covered with this ancient type of rock, which is believed to be the material of the original lunar crust. This sample's dimensions are 18 cm $\times$ 16 cm $\times$ 7 cm (7 in $\times$ 6 in $\times$ 3 in). Although this sample is medium gray, other anorthosites retrieved from the Moon have been white, while others are darker gray than this one. This rock was brought back by *Apollo 16* astronauts. (NASA)



FIGURE 6-26 Impact Breccias These rocks are created from shattered debris fused together under high temperature and pressure. Such conditions prevail immediately following impacts of space debris on the Moon's surface. (NASA)

(**Figure 6-26**), which are composites of different rock fused together as a result of meteorite impacts.

The rough, powdered regolith, composed of intrinsically dark basaltic and anorthositic rock, absorbs most of the light incident on it. The Moon's albedo is therefore only 0.12, meaning that it scatters only 12% of the light striking it.

Four of the six Apollo missions that landed on the Moon set up scientific packages containing seismic detectors (**Figure 6-27**). These devices are sensitive to moonquakes, motions of the Moon's surface analogous to earthquakes. Between 1969 and 1977, they detected more than 12,500 events, including a few quakes strong enough to knock books off shelves (if there had been shelves or books there at the time). Moonquakes occur for any of four reasons: impacts, motion of the lunar surface due to changing tides created by Earth's gravity, expansion and contraction due to heating and cooling of the Moon's surface, and other activity inside the Moon that is still under investigation.

With several seismic detectors working simultaneously, geologists were able to deduce information about the Moon's interior from the timing and strength of the vibrations that were detected. The data suggested that the Moon has a small iron core, possibly a layer of hot rock above it, and a thick layer of solid rock above that, surrounded by the crust. The presence of the iron core, accounting for only 1% to 4% of the Moon's total mass, was confirmed by data from the *Lunar Prospector* satellite in 1998. This is a small proportion compared to Earth's core, which has about 32.5% of our planet's total mass.

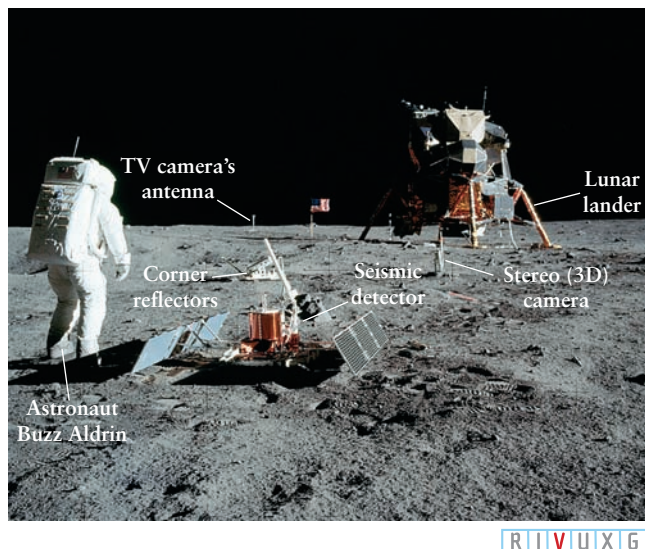


FIGURE 6-27 Apollo 11 Landing Site On the Moon's Sea of Tranquility, astronaut Buzz Aldrin stands next to the package of equipment containing the seismic detector. The corner reflectors are used, even today, to determine the distance from Earth to the Moon. The stereo camera took pairs of images of the Moon's surface. Seeing them through special glasses gives a 3-D close-up view of the Moon's surface. (Figure 6-23b is one of such a pair.) The bottom half of the lander is still on the Moon's surface. The top half brought astronauts Neil Armstrong and Buzz Aldrin back into lunar orbit, where they transferred to the command module to fly home. (NASA)

Like Earth, the Moon's interior was originally molten. Because the Moon is smaller than Earth, it initially had less heat stored inside. Heat from inside both worlds leaks out through their surfaces, with the smaller Moon cooling faster than Earth. This is why less, if any, of the Moon's interior is molten today.

Exposed to the cold of space, the Moon's surface solidified before its hot interior did. As the Moon's interior cooled and solidified, it became more compact; virtually all substances (except water) shrink when they become solid. Therefore, as the Moon's interior solidified, gravity pulled its already solid surface inward. This process could not happen uniformly (think of an apple that is drying on the inside). Therefore, parts of the Moon settled inward faster than other parts. As a result, sharp cliffs called **scarps** formed where one region of the surface moved down before adjacent regions could (Figure 6-28a). In 2010, astronomers using a camera on the *Lunar Reconnaissance Orbiter* observed the existence of scarps over much of the Moon's surface. Calculations indicate that the Moon has shrunk about 100 m (110 yd) as a result of its cooling.

While scarps are indicative of a shrinking world, in 2012 astronomers using the *Lunar Reconnaissance Orbiter* observed evidence that the Moon has regions where its surface is expanding.

The evidence is in the discovery of *graben* (Figure 6-28b), which form when nearby regions of a surface separate, allowing the intervening land to descend. The graben look like long troughs and may have formed as recently as 50 million years ago. Depending on how recently they formed, the graben could indicate that the inside of the Moon may still have some magma that is causing the surface to move. Hence, it appears that the Moon is contracting in some places and expanding in others!

Since spacecraft began orbiting the Moon, scientists have noted that their orbits do not follow the elliptical paths predicted by Kepler's laws. Every once in a while, usually over a mare, the spacecraft momentarily dip Moonward, as they are pulled by a higher-than-normal concentration of mass. These *mass concentrations*, or **mascons**, are local regions of relatively dense rock and metal near the Moon's surface. The mascons found in maria are believed to be the magma that filled the bottom of the mare after the basins were formed by impacts of very large bodies during the Late Heavy Bombardment. The magma solidified after emerging from inside the Moon. Before the mascons were well understood, a small satellite put in orbit in 1972 by *Apollo 16* astronauts came under their influence, and their stronger-than-average gravitational tug caused the satellite to crash into the Moon. The satellite, intended to orbit for years, lasted only 35 days. New mascons continue to be discovered, including seven found by the *Lunar Prospector* in 1998.

In 1994, the lunar-orbiting *Clementine* spacecraft sent radio signals to the Moon's surface. These signals scattered off the materials there and then were received and analyzed on Earth. The results were intriguing, suggesting that frozen water exists in craters near the Moon's southern pole. The presence of ice on the Moon was confirmed by the Lunar Crater Observing and Sensing Satellite (LCROSS), the Indian *Chandrayaan-1* lunar orbiter, and the *Lunar Prospector* orbiter. These craters alone, ranging in size from 2 to 15 km (1 to 9 mi) across, contain an estimated 600 million metric tons (670 million tons) of water ice. This ice, in polar craters that are never exposed to sunlight, is not in danger of being evaporated (Figure 6-28c). The ice was probably deposited on the Moon during comet and asteroid impacts.

While water ice at the poles, blocked from the Sun's heat by crater walls, does not surprise astronomers and geologists, the discovery by *Chandrayaan-1* in 2010 of water (in the form of OH^- , called the hydroxyl radical, and H, hydrogen atoms) in the Moon's equatorial regolith was unexpected. Even so, compared to Earth, the Moon is bone dry. Assuming that such water-bearing rock is distributed over most of the Moon's surface, it is estimated that the Moon contains as much water as is in the Great Lakes of North America, which have a volume of about $22,500 \text{ km}^3$ (5400 mi^3).

The presence of water ice would greatly simplify human settlement of the Moon, because ice can provide us not only with liquid water but also with rocket fuel and

Focus Question 6-7

What type of rock did astronauts bring back from *Apollo 11*, the first flight to land on the Moon?

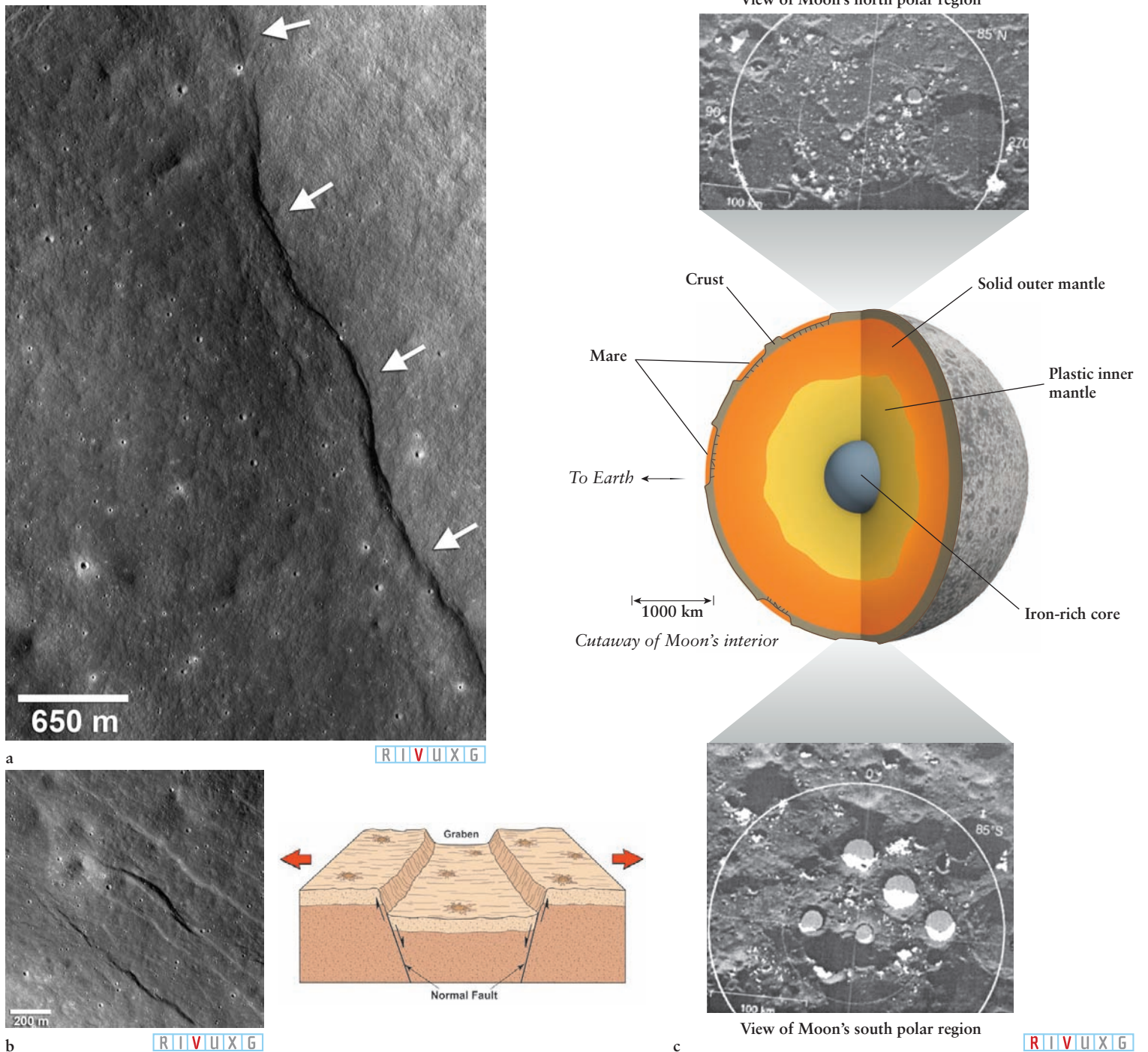


FIGURE 6-28 The Moon's Interior and Its Effects on the Moon's Surface (a) As the Moon's interior shrank, the surface settled irregularly, creating long lines of cliffs called scarps. (b) Image from the *Lunar Reconnaissance Observer* showing graben that are about 20 m (22 yd) deep. Inset shows the dynamics that created the graben. (c) Based on seismic experiments left on the Moon by Apollo astronauts and the studies done by *Lunar Prospector*, we know that the Moon has a crust, a mantle, and a core. The lunar crust ranges in thickness between about 20 and 120 km (12 and 75 mi), with an average thickness of about 60 km (37 mi) on its Earth-facing side and about 72 km (45 mi) on the far side. The crust and solid upper mantle extend inward to about 800 km (500 mi), where the nonrigid inner mantle begins. The Moon's core has a radius of

somewhere between 220 and 450 km (137 and 280 mi). Although the main features of the Moon's interior are analogous to those of Earth, the proportions and details are quite different. The *Clementine* spacecraft revealed that the south polar region on the far side of the Moon has a significant basin where the crust was apparently stripped away by an impact. **Top inset:** A radar image of the Moon's north polar region. The areas computer-colored in white and light gray are regions where the Sun never shines and that contain water. **Bottom inset:** A radar image of the Moon's south polar region, also showing regions of permanent or near-permanent darkness. (a: NASA/Goddard/Arizona State University/Smithsonian; b: NASA/Galaxy; b insets: Arizona State University/Smithsonian Institution; c: NASA/Galaxy)

breathable oxygen. The space shuttle's primary rocket engine was fueled by liquid oxygen and liquid hydrogen. Separating the Moon's water into liquid oxygen and liquid hydrogen would create the same fuel. An existing water reservoir would make establishing a lunar colony much less expensive and more feasible than lugging all that water from Earth or dragging a water-rich comet to the Moon.

Astronauts found no traces of life on the Moon. Life as we know it requires liquid water, of which the Moon has none, to form and evolve. The Moon has an atmosphere of helium, argon, sodium, and potassium that is so thin as to be considered an excellent vacuum compared to the density of the air we breathe. An atmosphere is held around a planet or moon by that body's gravitational force. If the atmospheric gases are too hot (moving too fast) for the gravitational force to hold them, they will leak into space. Our Moon has so little gravitational attraction (about one-sixth of what we feel at Earth's surface) that it is unable to retain its tenuous atmosphere. The atmosphere is continually being renewed by sunlight striking the Moon, releasing more gas from its rocks. Earth, in contrast, has enough mass to retain such gases as nitrogen, oxygen, carbon dioxide, water, and argon.

The Moon's slow rotation rate and small iron core suggest that, unlike Earth, the Moon should not have a strong magnetic field. Spacecraft orbiting the Moon have confirmed this. However, the *Lunar Prospector* has discovered local areas where the magnetic field is strong enough to keep the solar wind away from the Moon's surface. These anomalous regions have yet to be explained.

6-8 The Moon probably formed from debris cast into space when a huge planetesimal struck the young Earth

As noted in Section 4-4, most astronomers believe that the Moon formed when the newly formed Earth was struck at an angle by a Mars-sized planetesimal that literally splashed some of Earth's surface and some of the impactor's mass into orbit around the young planet. This is called the **collision-ejection theory** or *large impact theory*. This orbiting matter became a short-lived ring that eventually condensed to become the Moon, just as planetesimals condensed to form Earth. Evidence from Moon rocks places this event as occurring between 60 and 100 million years after the solar system began forming nearly 4.6 billion years ago. Earth had differentiated (see Section 6-4) before these times, so the material ejected from Earth that helped form the Moon came from Earth's crust and mantle but not its iron-rich core. A computer simulation of this cataclysm is shown in [Figure 6-29](#).

The collision-ejection theory is consistent with the known facts about the Moon. For example, rock

pulverized by the impact would have been depleted of *volatile* (easily evaporated) elements and water, leaving the Moon relatively low in water (compared to Earth), as we now know to be true. Also, the Moon has little iron-rich matter because that material had sunk deep into Earth before the planetesimal struck. The material from Earth that splashed into orbit and helped form the Moon was mostly the lighter rock floating on Earth's surface. Furthermore, most of the debris from the collision would orbit near the plane of the ecliptic as long as the orbit of the impacting planetesimal had been near that plane. Also, the impact of an object large enough to create the Moon could have tipped Earth's axis of rotation, inaugurating the seasons ([Figure 6-30](#)). The fact that this theory is both consistent with observations and explains two seemingly separate things (the existence of the Moon and the tilt of Earth) makes it especially appealing to scientists.

The surface of the newborn Moon probably remained molten for millennia, due to heat released during the impact of rock fragments falling onto the young satellite and the decay of short-lived radioactive isotopes inside it. As the Moon gradually cooled, low-density lava floating on its surface began to solidify into the anorthositic crust that exists today in the highlands.

By correlating the ages of Moon rocks with the density of craters at the sites where they were collected, geologists have determined how the rate of impacts on the Moon changed over the ages. In summary, the ancient, heavily cratered lunar highlands are evidence of an intense bombardment of asteroids that dominated the Moon's early history. This barrage began 4.5 billion years ago, when the Moon's surface first solidified, and ended about 3.8 billion years ago. This period of impacts was punctuated by a severe pounding 4 billion years ago, the Late Heavy Bombardment, which lasted for millions of years. Recall from Chapter 4 that, according to the Nice model, this bombardment was caused by the migration of the giant planets.

The frequency of impacts gradually tapered off as meteoroids and planetesimals were swept up by the newly formed planets, as discussed in Chapter 4. Recorded among the final scars at the end of this crater-making era are the impacts of more than a dozen objects, each measuring at least 100 km (62 mi) across. As these huge rocks slammed into the young Moon, they blasted gigantic craters. Meanwhile, heat from the decay of long-lived radioactive elements, like uranium and thorium, began to remelt the inside of the Moon. Then, from 3.8 to 3.1 billion years ago, great floods of molten rock from just below the surface (not from the Moon's core) gushed up from the lunar interior through the weak spots in the crust created by these large impacts. Lava filled the impact basins and created the maria we see today. Because most of the bombardment was over by then, maria have few craters.

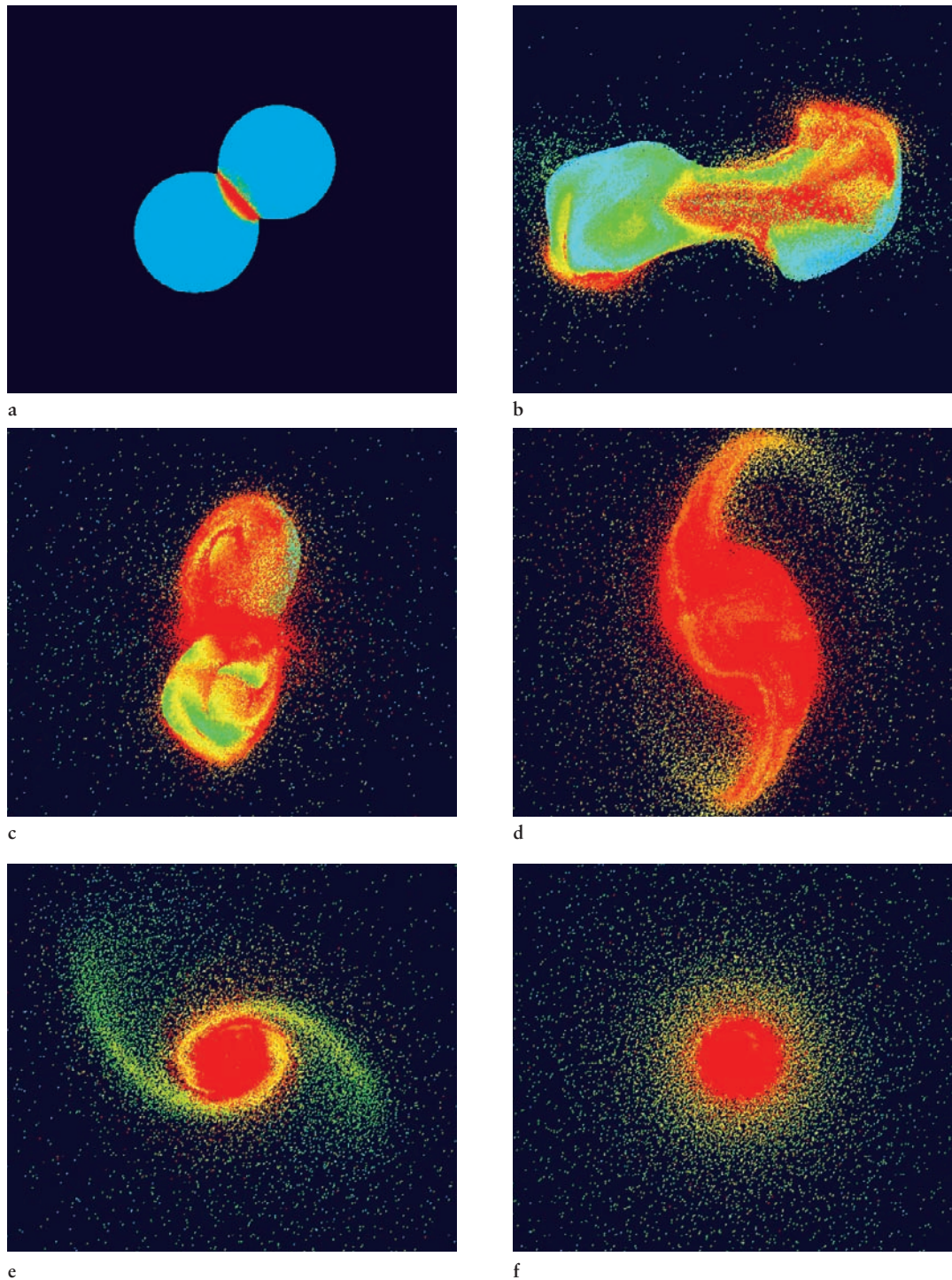


FIGURE 6-29 The Moon's Creation This computer simulation models the creation of the Moon from material ejected by the impact of a large planetlike body with the young Earth. Successive views (**a–f**) show the impact, splash, and formation of a ring of debris orbiting Earth. In this simulation, the impact

ejects both mantle and core material, but most of the dense iron falls back onto Earth. The surviving ejected rocky matter forms the ring in frame (f). This debris will quickly coalesce into the Moon, initially orbiting some 10 times closer to Earth than it is today. (Courtesy of Robin Canup)

The reason that the far side of the Moon has only one significant mare apparently stems from the Moon's crust on the side facing Earth being thinner than the

crust on the far side, as noted in Section 6-6. This difference was created when the Moon was young and molten. At that time, the Moon was much closer to Earth

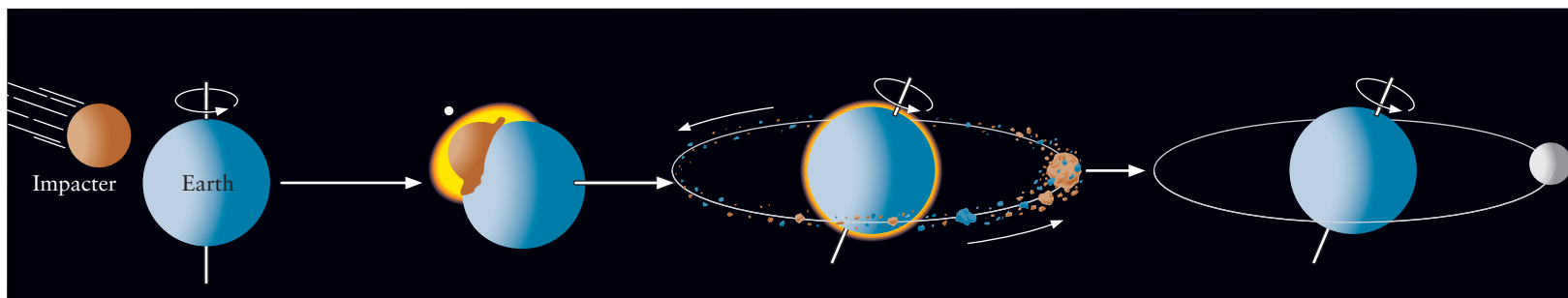


FIGURE 6-30 Tilting Earth's Axis The collision that created the Moon could have also knocked Earth's rotation axis over so that today it has a $23\frac{1}{2}^\circ$ tilt, thereby creating the seasons.

(we will explore this shortly). The Moon's rapid orbit and the molten rock tides created on it by Earth caused much of the low-density, crustal lava to flow away from Earth to the lunar far side, where it remained when the crust solidified. It was therefore easier for the molten rock inside the Moon to escape through the large craters on the thin crust of our side than through those on the thicker-crust far side. It is worth noting a recent proposal that the near and far sides are so different because Earth may have once had two moons that spiraled into each other. Details of this event can be found in the first chapter of the book *What If the Earth Had Two Moons?*

Relatively little large-scale activity has happened on the Moon since those ancient times. Although only a few fresh large craters have been formed, the small-sized debris still plentiful in the solar system continues to create small craters. Back in 1953, an impact was observed that has been connected with the formation of a crater

having a diameter of 1.5 km (0.93 mi).

Smaller impacts are observed nearly every year when large groups of rocky space debris, left in orbit from comets as they vaporize, cross the Moon's path. By and large, however, the astronauts visited a world that has remained

largely unchanged for more than 3 billion years. Unlike Earth, which is still quite geologically active, any molten rock inside the Moon has apparently stopped leaking onto that world's surface. Astronomers hope to learn more about the Moon's interior when modern seismographs are eventually placed on its surface.

6-9 Tides have played several important roles in the history of Earth and the Moon

Gravitational attraction keeps Earth and the Moon together, while the angular momentum (see Appendix P: Energy and Momentum) given to the Moon when it was formed keeps the Moon from being pulled back onto Earth. Contrary to appearances, the Moon does not orbit Earth. Rather, both bodies are in motion around their common center of mass, called their *barycenter*

(Figure 6-31a), just as a pair of objects tied together and sent spinning across an air table move around a similar common point (Figure 6-31b). This motion is crucial in understanding the cause of the tides, but before explaining them, it is instructive to introduce another tide-related issue, namely, that the rate at which the Moon orbits the barycenter and the rate at which it rotates are the same.

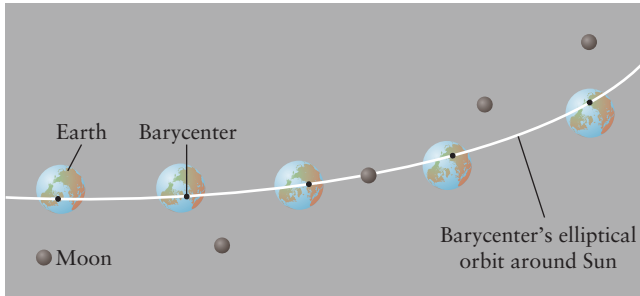
2 3 Except for slight variations as described below, the same surface features on the Moon always face Earth, regardless of what phase the Moon is in (see Figure 1-20). This phenomenon occurs because the Moon is in **synchronous rotation** around Earth: The Moon rotates on its axis once every 27.3 days, the same length of time it takes the Moon and Earth to orbit their barycenter (Figure 6-32). To see why the Moon must be rotating if it keeps the same side facing us, imagine standing up, holding your arm out in front of you with your palm facing you, and twirling yourself around. You are Earth, while your palm represents the side of the Moon facing Earth. The only way your palm can always face you (be in synchronous rotation) is if your hand (the Moon) rotates at exactly the same rate that it swings (revolves). So the Moon must be rotating. It is interesting to note that most moons in our solar system are in synchronous rotation around the planets they orbit. We will discuss their orbits further in Chapter 7.

We can see more than half the Moon's surface from Earth even though it is in synchronous rotation with respect to our planet. This occurs for several reasons; namely, the Moon's orbit is not perfectly circular; its rotation axis is not perpendicular to the plane of its orbit around Earth; the plane in which it orbits is tilted relative to the ecliptic; and we can observe it from different places on the Earth's surface. As a result, we actually see a little beyond the poles and around the sides of the Moon. We can see a total of about 59% of the Moon's total surface from Earth at different times throughout the year. The changes in what we see of the Moon from Earth are called *librations*.

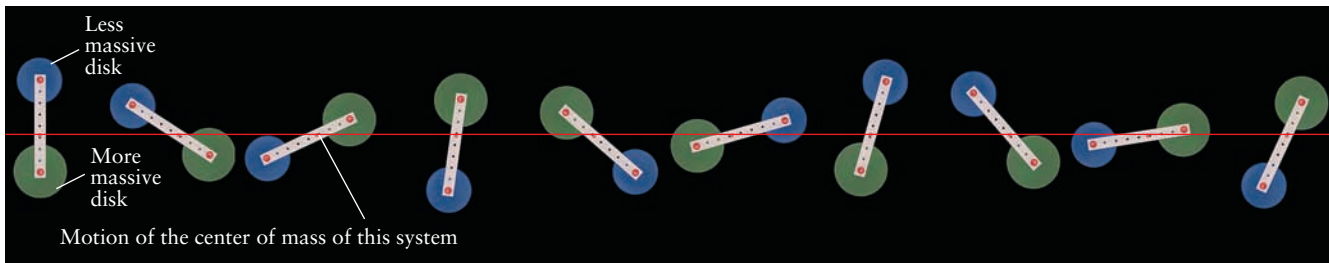
We have seen the Moon's complete "far side" only from spacecraft (see Figure 6-21). The far side of the

Focus Question 6-8

Why are we fortunate on Earth that the Moon is rarely being struck today?



a



b

Moon is often, and incorrectly, called its “dark side.” The dark side of the Moon is exactly like the dark side of Earth—it is the side where the Sun is below the horizon. Whenever we view less than a full Moon, we see part of its dark side, while at the same time some sunlight is then striking part of its far side.

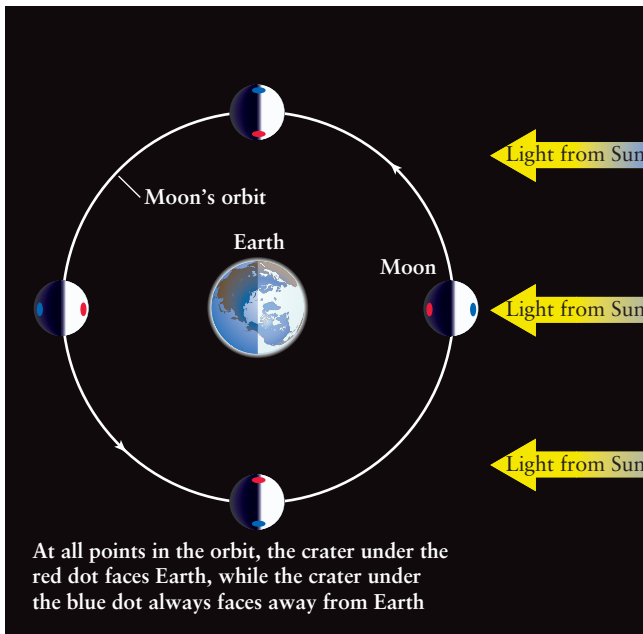


FIGURE 6-32 Synchronous Rotation of the Moon The motion of the Moon around Earth as seen from above Earth's north polar region (ignoring Earth's orbit around the Earth-Moon barycenter). For the Moon to keep the same side facing Earth as it orbits our planet, the Moon must rotate on its axis at the same rate that it revolves around Earth.

FIGURE 6-31 Motion of Earth-Moon System (a) The paths of Earth and the Moon as their barycenter follows an elliptical orbit around the Sun. (b) Analogously, this time-lapse image shows a pair of different-mass disks connected by a piece of wood sliding across a table. Note that the center of mass of this collection of objects moves in a straight line, while its other parts follow curved paths. As with the Earth-Moon system, the center of mass is closer to the more massive disk than to the less massive one.

4 The origin of the Moon's synchronous rotation is the tidal force from Earth that acted on it shortly after the Moon formed. At the same time, the tidal force from the Moon began creating ocean tides here. To understand both how the Moon came to have synchronous rotation and our present tides, we investigate the forces acting between Earth and the Moon. The discussion in Appendix O: Tides focuses on how the Moon creates the tides on Earth.

Two-thirds of the tide height on Earth today results from the gravitational tug-of-war between Earth and the Moon as the two bodies orbit their barycenter. The Sun generates most of the remaining third of the tides, with Jupiter and the other planets contributing a tiny fraction.

As explained in Appendix O: Tides, two high tides always occur on opposite sides of Earth (see Figure O-1). When the Sun, the Moon, and Earth are aligned, the Sun and the Moon create pairs of high ocean tides in the same directions, and the resulting combined tides are the highest high tides (and lowest low tides) of the lunar cycle. These **spring tides** (so called not because they only occur in the spring—they occur in all seasons—but because the term derives from the German word *springen*, meaning to “spring up”) occur at every new and full Moon (see Figure O-2). Note that it does not matter whether the Sun and Moon are on the same or opposite sides of Earth when generating the spring tides, because the Sun and the Moon both create high tides on opposite sides of Earth.

At first quarter and third quarter, the Sun and the Moon form a right angle as seen from Earth. The gravitational forces they exert compete with each other, so

the tidal distortion is the least pronounced. The especially small tidal shifts on these days are called **neap tides**.

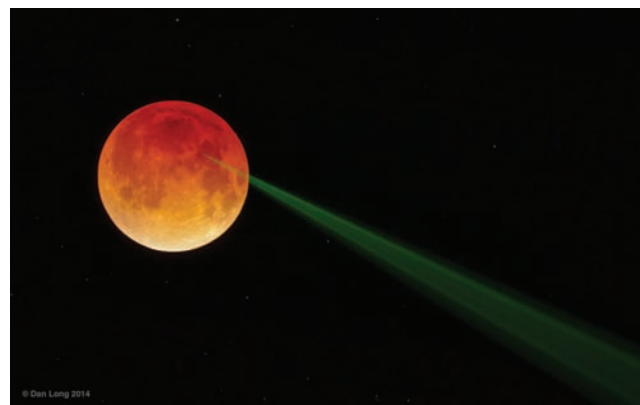
Finally, we can explain the Moon's synchronous rotation. When the Moon was young, it was molten and Earth's gravity created huge tides of molten rock up to 18 m (60 ft) high there. Friction generated by these tides between the liquid rock and the mantle rock below it caused the Moon to speed up or slow down (we do not know which) until one of the high tides became fixed directly between the Moon's center and the center of Earth. As the surface solidified, it locked onto the interior, causing the entire Moon to rotate at the same rate as its surface. The entire Moon was then in synchronous rotation.

6-10 The Moon is moving away from Earth

In 1897, Sir George Darwin, son of the evolutionist Charles Darwin, proposed that the Moon is moving away from Earth. To understand why, read Appendix O: Tides and see Figure O-2, which shows that the high tide nearest the Moon is actually 10° ahead of the Moon in its orbit around Earth. This offset occurs because friction between the ocean and the ocean floor causes the rapidly spinning Earth to drag the tidal bulge ahead of the line between the center of Earth and the center of the Moon. The gravitational force from the water in the high ocean tide nearest the Moon acts back on the Moon, pulling it ahead in its orbit, giving it energy and thereby forcing it to spiral outward. This effect is similar to what would happen if you tied a ball on a string, let it hang down, and then started spinning yourself around. Your hand would act through the string, giving the ball energy and starting it to spin and move farther and farther away from your body. Contrary to intuition, however, giving the Moon energy to spiral outward does not mean that it is continually moving faster. In fact, the farther it moves away from Earth, the slower it orbits the barycenter.

To test the concept that the Moon is moving away from Earth, *Apollo 11*, *14*, and *15* astronauts placed sets of reflectors on the Moon (see Figure 6-27), similar to the orange and red ones found on cars. These reflectors are specially designed so that light striking them from any direction is reflected back toward its source. Pulses of laser light were fired at the Moon from Earth (Figure 6-33), and the time it took the light to reach the Moon, bounce off a reflector array, and return to Earth was carefully recorded. Knowing the speed of light and the time it takes for the light

pulse to make the round trip, astronomers can calculate the distance to the Moon with an error of only a few millimeters. From such measurements made since 1969, astronomers have established that the Moon is currently spiraling away from Earth at a rate of 3.8 cm (1.5 in) per year.



RITVUXG

FIGURE 6-33 Lunar Ranging Beams of laser light are fired through a telescope at the Apache Point Observatory in New Mexico. The light is then reflected back by the corner reflectors placed on the Moon by *Apollo 15* astronauts. From the time it takes the light to reach the Moon and return to Earth, astronomers can determine the distance to the Moon to within a few millimeters. (Dan Long, Apache Point Observatory)

Since it is moving away, the Moon was clearly closer to the Earth in the past, as discussed in Appendix O: Tides. Simulations suggest that it was perhaps as much as 10 times closer when it first formed (as indicated in Figure 6-29). Contrary to intuition, however, the Moon will never leave Earth's orbit. The Earth's rotation rate slows as the Moon recedes (see Appendix O). Eventually the Earth will rotate at the same rate the Moon orbits, thereby forcing the closer high tide to be directly under the Moon. Lacking the forward tug from this tide that it receives today, the Moon will stop spiraling outward.

MERCURY

The closest planet to the Sun, Mercury (Figure 6-34) is a truly inhospitable world of temperature extremes. Its incredibly thin atmosphere and weak magnetic field leave it virtually unprotected from countless impacts and a continuous bath of deadly solar radiation. Despite having a surface nearly as dark as coal, Mercury sometimes appears as one of the brightest objects in our sky, due to its proximity to the Sun. We can see it from Earth only for a few hours before sunrise or a few hours after sunset because its angle from the Sun (its elongation) is always less than 28° . Because its orbit around the Sun is so elliptical, Mercury's maximum elongation (see Figure 2-8c) ranges between 18° and 28° .

6-11 Photographs from *Mariner 10* and *Messenger* spacecraft reveal Mercury's lunarlike surface

Until late last century, Mercury was a complete mystery. Earth-based telescopic studies provided virtually no

Focus Question 6-9

Why haven't the ocean tides on Earth put our planet into synchronous rotation with respect to the Moon?

MERCURY: Vital Statistics

Planet symbol:	♿
Average distance from the Sun:	0.387 AU = 5.79×10^7 km = 3.60×10^7 mi
Maximum distance from the Sun:	0.467 AU = 6.98×10^7 km = 4.34×10^7 mi
Minimum distance from the Sun:	0.307 AU = 4.60×10^7 km = 2.86×10^7 mi
Orbital eccentricity:	0.21
Average orbital speed:	47.9 km/s
Sidereal period of revolution (year):	88.0 Earth days = 0.24 Earth year
Sidereal rotation period:	58.6 Earth days
Solar rotation period (day):	176 Earth days
Inclination of equator to orbit:	0.5°
Inclination of orbit to ecliptic:	7° 00' 16"
Diameter (equatorial):	4880 km = 0.382 Earth diameter
Mass:	3.30×10^{23} kg = 7.3×10^{23} lb = 0.0553 $M_{\oplus}$
Average density:	5430 kg/m ³ = 9150 lb/yd ³ = 0.984 Earth density
Escape speed:	4.3 km/s = 2.7 mi/s
Surface gravity (Earth = 1):	0.38
Albedo:	0.12
Average surface temperatures:	Day: 350°C = 662°F = 623 K Night: -170°C = -274°F = 103 K
Surface air pressure:	10^{-14} × Earth's surface air pressure
Atmosphere:	Transient hydrogen (H ₂), helium (He), potassium (K), sodium (Na), and oxygen (O ₂)

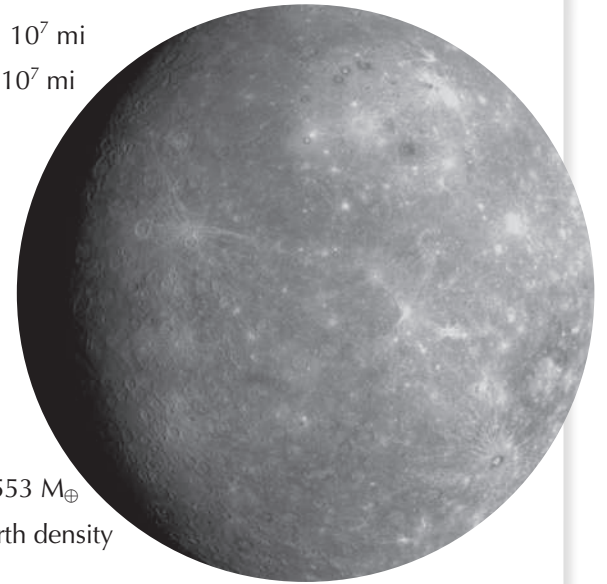


FIGURE 6-34 Mercury's Vital Statistics NASA's *Mariner 10* spacecraft flew past heavily cratered Mercury 3 times in 1974 and 1975. The *Messenger* spacecraft has been orbiting Mercury

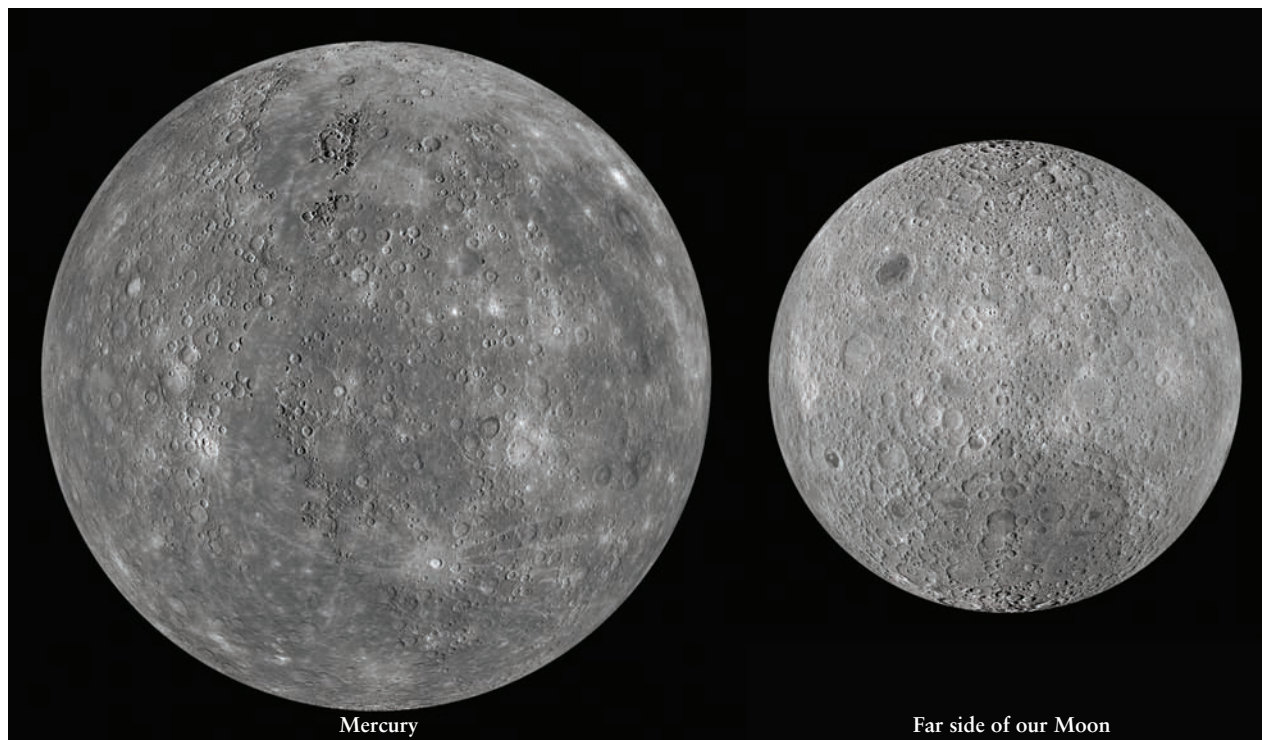
since March 18, 2011. This image was taken by *Messenger* from an altitude of 27,000 km (17,000 mi). (NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington)

information about the planet's surface. The first spacecraft to visit it, *Mariner 10*, went by on three separate occasions, between March 1974 and March 1975, but only photographed 45% of Mercury's surface. The images were tantalizing, showing the planet to be heavily cratered. The only other spacecraft to visit it, NASA's *Messenger*, was put into orbit in 2011 and has imaged the rest of the planet. *Messenger* is short for *M*ercury *S*urface, *S*pace, *E*Nvironment, *G*Eochemistry, and *R*anging. It is also reminiscent of the historical Roman assignment of Mercury as the messenger of the gods.

First impressions of Mercury evoke a lunar landscape, and there are several similarities. Craters in the Moon's highlands are densely packed, as are some of the craters on Mercury (Figure 6-35). Astronomers conclude that most of the craters on both Mercury and the Moon were produced by impacts in the first 800 million years after these bodies condensed from the solar nebula.

The impact craters on both worlds have similar features, such as ejecta blankets and central peaks, as discussed for the Moon in Section 6-6. Unlike our Moon, however, Mercury also has a variety of volcanic craters out of which poured large amounts of lava that filled smaller impact craters and created smooth plains between larger ones (Figure 6-36). This volcanic activity may explain the high concentrations of sulfur-based compounds that *Messenger* has discovered on the planet's surface. Like the Moon, Mercury has an exceptionally low albedo of 0.12 (its surface scatters 12% of incoming light). It is very bright as seen from Earth only because the sunlight scattering from Mercury is so intense.

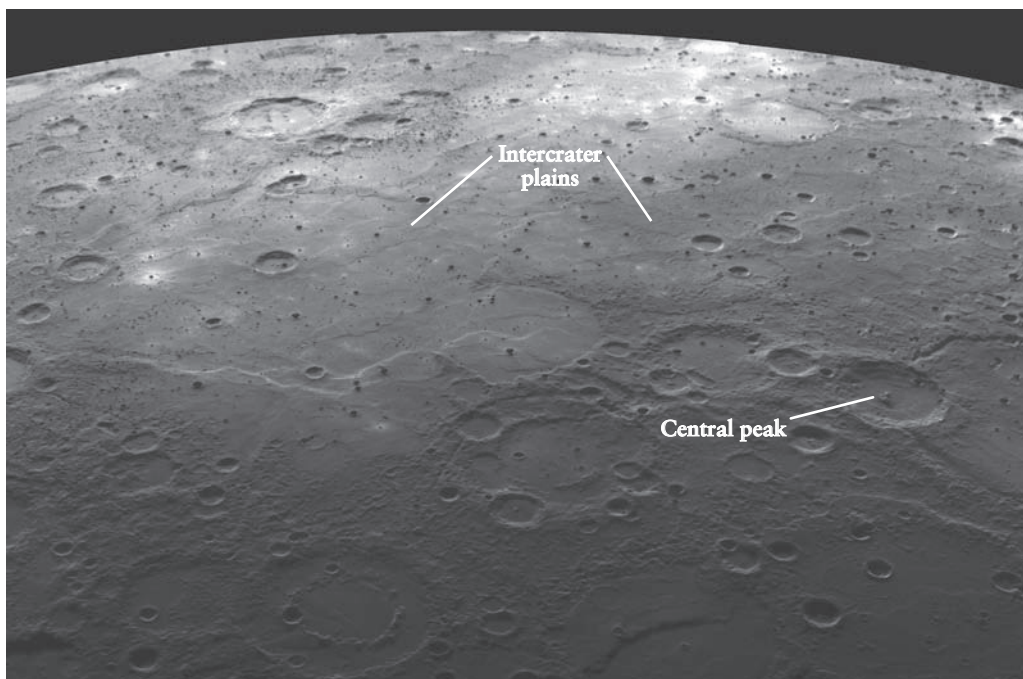
The most impressive feature discovered by *Mariner 10* is a huge circular region called Caloris Basin. This name derives from *caloris*, Latin for "heat," and in fact this feature is the closest part of Mercury to the Sun at the planet's perihelion (closest approach to the Sun).



R I V U X G

FIGURE 6-35 Mercury and the Far Side of Our Moon Mercury and our Moon are shown here to the same size scale. Mercury's radius is 2439 km (1515 mi) and the Moon's is 1738 km (1080 mi). For comparison, the distance from New York to Los Angeles is 3944 km (2451 mi). Mercury's surface

is more uniformly cratered than that of the Moon. Daytime temperatures at the equator on Mercury reach 700 K (800°F), hot enough to melt lead or tin. (a: NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington b: NASA/GSFC/Arizona State University)



R I V U X G

FIGURE 6-36 Mercury's Craters and Plains This view of Mercury's northern hemisphere was taken by *Messenger* as it sped past the planet in 2009. Numerous craters on the top left and bottom half of the image are separated by a broad intercrater plain created by lava flow from now extinct volcanoes. The image is about 880 km (550 mi) across. (NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington)

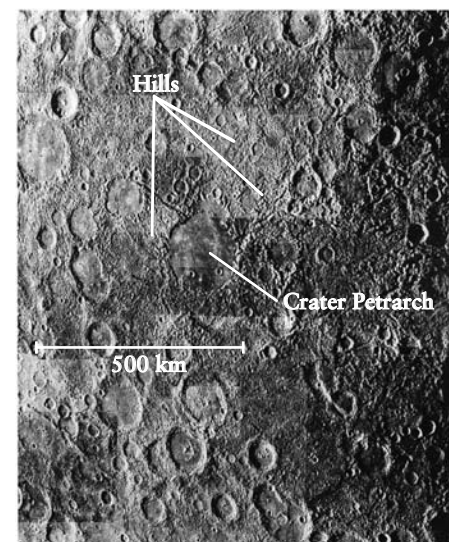
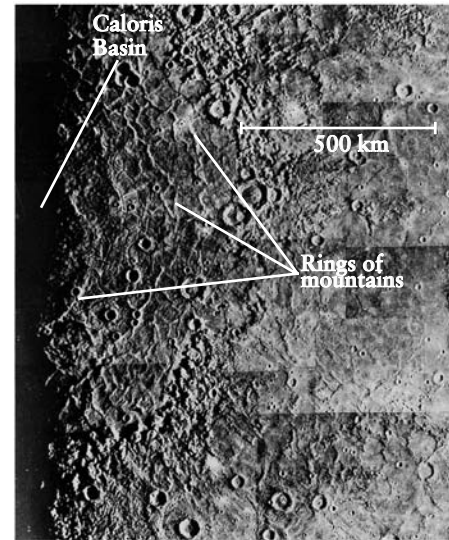
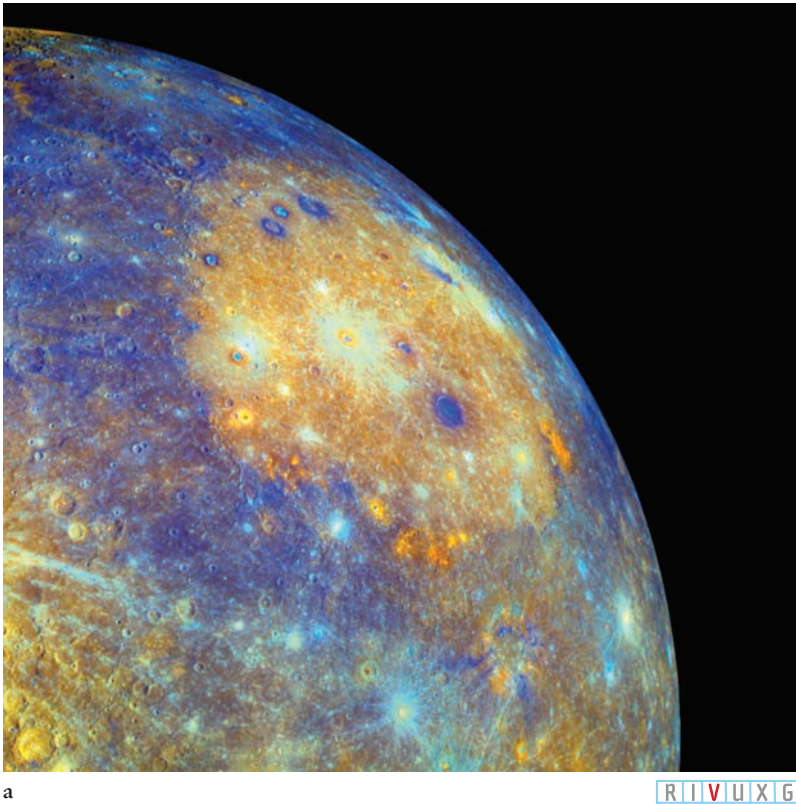


FIGURE 6-37 Major Impacts on Mercury (a) **Caloris Basin.** *Messenger* sent back this view of a huge impact basin on Mercury's equator. The basin is the entire orange-colored portion of the figure. This image has been color enhanced to show the different surface compositions of Mercury. The orange regions around the edge of the basin are believed to be volcanic features. (b) Only about half the Caloris Basin appears here because it happened to lie on the terminator when the *Mariner 10* spacecraft sped past the planet. Although the center of the impact basin is hidden in the shadows (just beyond the left side of the picture), several semicircular rings of mountains reveal its extent. (c) **Unusual, Hilly Terrain.** What looks like tiny, fine-grained wrinkles on this picture are actually closely spaced hills, part of a jumbled terrain that covers nearly 500,000 km² (193,000 mi²) on the opposite side of Mercury from the Caloris Basin. The large, smooth-floored crater, Petrarch, has a diameter of 170 km (106 mi). This impact crater was produced more recently than Caloris Basin. (a: NASA/APL; b and c: NASA)

Caloris Basin (Figure 6-37a) measures 1550 km (963 mi) in diameter, meaning that it has nearly 3 times the area of Texas. When viewed along the *terminator* (the border between day and night), where the shadows are longest and details clearest, the basin (Figure 6-37b) is seen to be surrounded by a 2-km-high (1.2-mi-high) ring of mountains, beyond which are relatively smooth plains. Like the lunar maria, Caloris Basin was probably gouged out by the impact of an asteroid that penetrated the planet's crust. *Messenger* revealed small impact craters inside the Caloris Basin that have been filled in by lava from volcanoes that were active long after both the basin and those craters were formed. A giant volcano 95 km (59 mi) in diameter (bigger than the state of Delaware) has been seen in the basin.

The Caloris impact was a tumultuous event that shook the entire planet. Indeed, the collision affected the side of Mercury directly opposite the Caloris Basin (Figure 6-37c), which has a jumbled, hilly surface covering nearly half a million square kilometers (190,000 mi²) or about twice the size of Wyoming. The hills, which appear as tiny wrinkles that cover most of the photograph, are about 5 to 10 km (3 to 6 mi) wide and between 100 and 1800 m (110 and 1970 yd) high. Geologists believe that energy from the Caloris impact traveled through the planet and became focused, like light through a lens, as it passed through Mercury. As this concentrated energy reached the far surface of the planet, jumbled hills were pushed up. Similar pairs of phenomena to Caloris and the jumbled terrain on the other side of Mercury can

be seen on our Moon. Both Mare Orientale and Mare Imbrium have chaotic hills on the opposite sides of the Moon from them. Finding such a similar pair of phenomena on a different world supports the connection between the impact and the jumbled surface.

Looking at Mercury in more detail, we start to see a variety of differences from the features on the Moon. Unlike our Moon's maria, Mercury lacks extensive craterless regions. Instead, Mercury has broad plains dotted with relatively small craters (see Figure 6-36). Such plains are not found on the Moon. As we learned in Section 6-6, the lunar maria were produced by extensive lava flows into giant preexisting craters, flows that occurred between 3.1 and 3.8 billion years ago. The reason Mercury has no maria is that it lacked the giant craters from which to form them.

Dormant volcanoes were discovered on Mercury in 2008 (Figure 6-38) by the *Messenger* spacecraft. Some of the material creating the Mercurian plains came from

Focus Question 6-10

The rings of mountains frozen into Mercury's and the Moon's surfaces have what *transient* analog on Earth? (*Hint*: Think about impacts.)

such volcanoes, while the rest came through craters created as large meteorites punctured the planet's thin, newly formed crust, allowing lava to well up from the molten interior to flood low-lying areas. The existence of craters on Mercury's plains suggests that these features formed just over 3.8 billion years ago, near the end of the Late Heavy Bombardment. Mercury's plains are therefore older than most of the lunar maria, leaving more time for cratering to occur on Mercury. Today, 40% of Mercury's surface is covered by these plains, while the rest is more heavily cratered.

Mariner 10 and *Messenger* also revealed numerous long cliffs, called scarps, meandering across Mercury's surface (Figure 6-39). Recall that they have also been observed on our Moon (see Figure 6-28). Just as occurred on our Moon (Section 6-7), scarps are believed to have developed as Mercury cooled, suggesting that the planet's interior is solid to a significant depth. Otherwise, lava would have leaked out as the scarps formed.

Messenger continues to provide intriguing images of Mercury. Figure 6-40 is one of the most interesting photos taken so far. The features displayed in this photograph include lines of secondary craters, created by ejecta from impacts; craters filled with volcanic lava; a large impact crater (diameter of 210 km [130 mi]) and concentric ring created as a ripple of surface rock by the impact, both of which are filled in with lava; a scarp; and a crater with as yet unexplained light rock revealed inside it. The large double-ringed crater is a smaller version of Caloris Basin.

6-12 Mercury has a higher percentage of iron than Earth

Mercury's average density of 5430 kg/m^3 (4.15 tons/yd³) is quite similar to Earth's (5520 kg/m^3 or 4.22 tons/yd³).

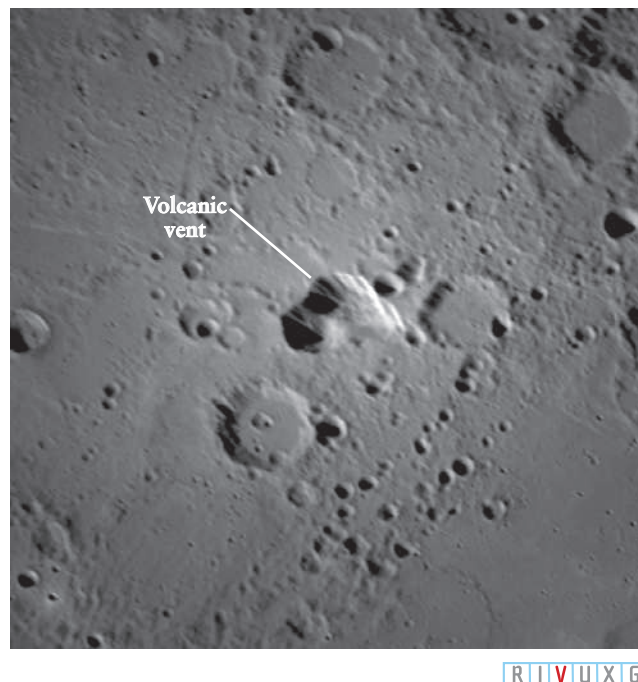


FIGURE 6-38 Possible Volcanic Vent The central indentation in this *Messenger* image from 2009 is believed to be a caldera (sunken vent) of an explosive volcano on Mercury. It is unlikely to be an impact crater, as it completely lacks a raised crater wall. (NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington)

As we saw in Section 6-4, typical rocks on Earth's surface have a density of only about 3000 kg/m^3 (2.3 tons/yd³) because they are composed primarily of lightweight

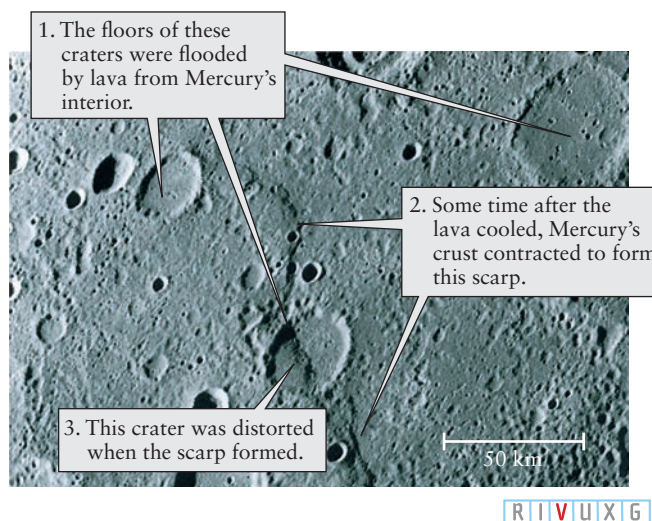
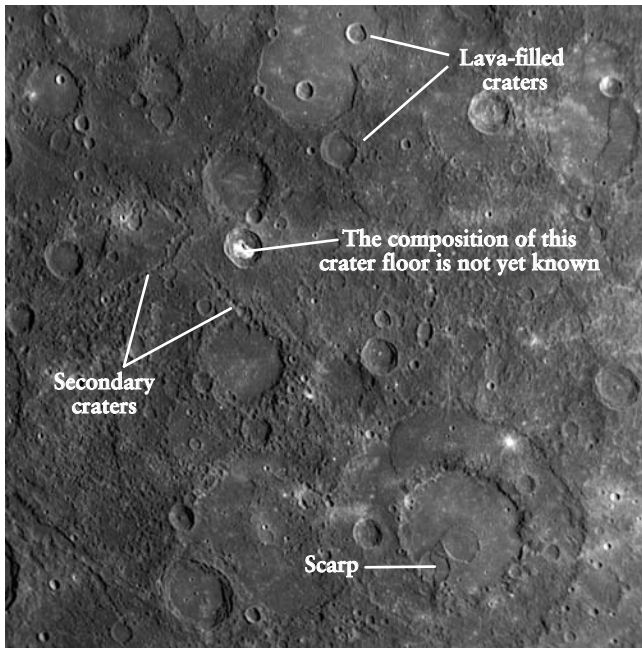


FIGURE 6-39 Scarps on Mercury A long, meandering cliff, called Santa Maria Rupes, runs from north to south across this *Mariner 10* image of a region near Mercury's equator. This cliff, called a *scarp* by geologists, is more than 1 km (0.6 mi) high and runs for several hundred kilometers. Note how the crater below the center of the image was distorted vertically when the scarp formed. (NASA)



R I V U X 6

FIGURE 6-40 Geology on Mercury This *Messenger* image shows a variety of interesting features, including lava-filled craters, secondary craters created by debris that splashed out of an impact crater, a scarp cutting through an old crater, and as yet unidentified, light-colored rock at the bottom of a crater. (NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington)

elements. The high average densities of both Mercury and our planet are caused by their dense interiors.

Because Mercury is less dense than Earth, you might conclude that Mercury has a lower percentage of iron, a common heavy element, than does our planet. Indeed, *Messenger* revealed that Mercury's surface has a low iron content. However, in terms of the percentage of its total mass that is iron, Mercury is the most iron-rich planet in the solar system. Only Earth's greater mass, pressing inward and thereby compressing our planet's inner parts, makes it denser than Mercury. **Figure 6-41** shows a scale drawing of Mercury's interior, where an iron core contains nearly 60% of the planet's mass and 75% of its radius. Surrounding the core is a rocky mantle 600 km (375 mi) thick. For comparison, Earth's iron core contains only 32.5% of Earth's mass and it extends out to only 55% of our planet's radius.

Mariner 10 discovered that Mercury has a varying magnetic field about 1% as strong as that of Earth. The field, aligned with Mercury's rotation axis, sometimes interacts with the solar wind (**Figure 6-42**) more weakly than does Earth's field, and sometimes more strongly (recall Section 6-5). Indeed, scientists have observed Mercury's field changing significantly over just a few years. The cause of the change is still unknown. Until 2008, astronomers did not even know what generated Mercury's magnetic field. In that year, the *Messenger*

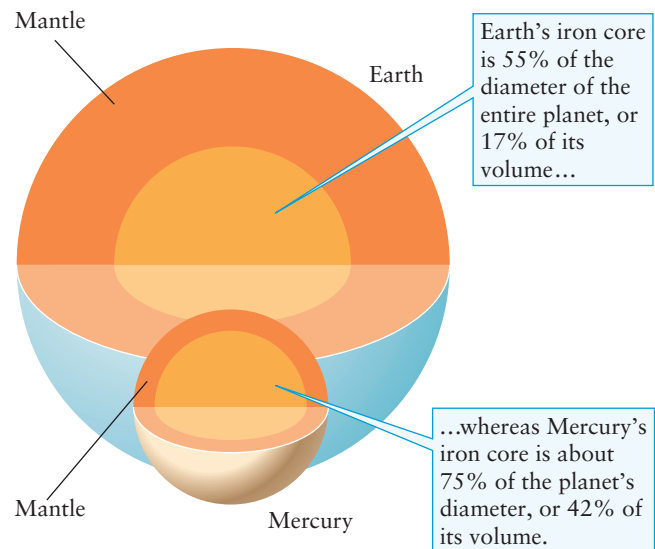


FIGURE 6-41 The Interiors of Earth and Mercury Mercury has the highest percentage of iron of any planet in the solar system. Consequently, its iron core occupies an exceptionally large fraction of its interior.

spacecraft passing by Mercury revealed that part of the planet's core is molten. As we saw in Section 6-5, electric currents in a planet's liquid iron core create a planetwide magnetic field. Earth creates its electric current by rotating once a day. Mercury rotates 59 times more slowly. That slower rotation and the lower volume of molten iron inside Mercury (compared to Earth) explain why Mercury's magnetic field is so weak.

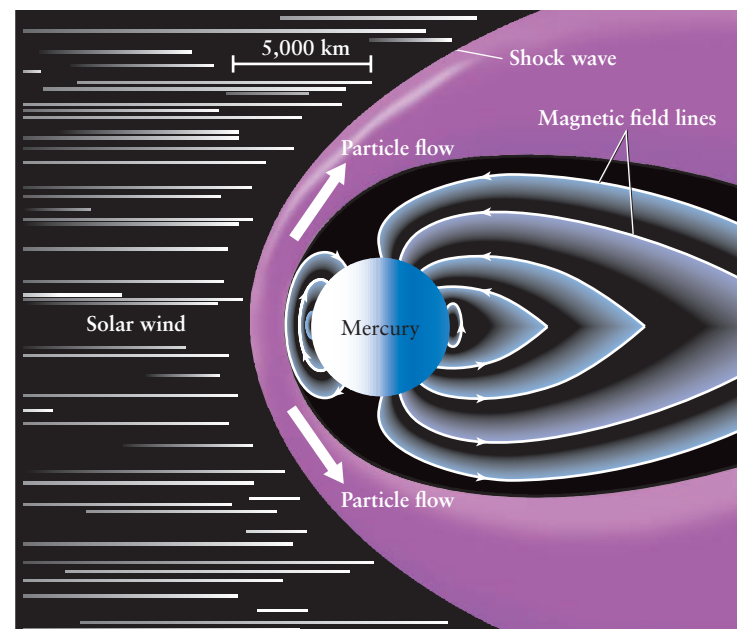


FIGURE 6-42 Mercury's Magnetosphere Mercury's magnetic field is about 1% as strong as Earth's field, just strong enough to deflect the solar wind.

Events early in Mercury's history must somehow account for its high iron content. We know that the inner regions of the primordial solar nebula were incredibly hot. Perhaps only iron-rich minerals were able to withstand the solar heat there, and these subsequently formed iron-rich Mercury. According to another theory, an especially intense outflow of particles from the young Sun stripped Mercury of its low-density mantle shortly after the Sun formed. A third possibility is that, during the final stages of planet formation, Mercury may have been struck by a large planetesimal (debris formed as the solar system came into being), just as Earth was struck by a Mars-sized body that led to the formation of our Moon. Computer simulations show that this cataclysmic collision would have ejected much of Mercury's lighter mantle (Figure 6-43). Such an impact could also help explain Mercury's significantly elliptical orbit; however, this theory is still being explored.

6-13 Mercury's rotation and revolution are coupled

Focus Question 6-11

If Mercury were struck by a large planetesimal, why would this collision not produce a moon, as happened when Earth was struck early in its history?

Mercury has one of the most unusual orbits in the solar system. Recall from Section 6-9 that when Earth was young, its gravitational force created tides on the Moon, thereby forcing the Moon into synchronous rotation. Something similar (but not identical) happened with the Sun playing the role of Earth and Mercury acting in the place of our Moon. To begin with, the Sun's gravity produced two significant tidal bulges on opposite sides of young, molten Mercury. Mercury is so close to the Sun (average separation 0.387 AU) that the Sun's gravitational force on Mercury's tidal bulges changed the planet's rotation rate before its surface

solidified, with the two regions of high tide now locked into place on the planet. However, Mercury's highly eccentric orbit ($e = 0.21$) prevented the planet from being locked into synchronous orbit like our Moon. Instead, Mercury developed what is called a **3-to-2 spin-orbit coupling**. This coupling means that Mercury undergoes three sidereal rotations (rotations measured with respect to the distant stars, not the Sun), while undergoing two revolutions around the Sun (Figure 6-44).

The major consequence of spin-orbit coupling is that one or the other of Mercury's regions of high tide is facing the Sun whenever the planet is at perihelion (see Figure 6-44).

A day on Mercury is 2 Mercury years long The motion of the Sun across Mercury's sky is unique in the solar system. First, a solar day there (noon to noon) is 176 Earth days long, twice the length of a year on Mercury! Furthermore, if you were to set up a camera at the location of high tide on Mercury to take a series of pictures, you would see the Sun start to rise in the east, stop high in the sky, move back toward the east, stop again, and then resume its westward journey. This is analogous to the retrograde motion of the planets that we observe from Earth (see Section 2-2).

Just as astronomers use scattered radio waves to search for water on the Moon (see Section 6-7), they have also sent radio waves to Mercury. In 1992, researchers made an extraordinary discovery—evidence for ice near Mercury's poles in craters that are permanently in shadow. Dozens of circular regions, now known to be craters, sent back signals with characteristics distinct to ice. The presence of this ice was confirmed by the *Messenger* spacecraft. The origin of the ice—whether from cometary impacts, from gases rising from inside the planet and then freezing, or from both sources—remains to be determined.

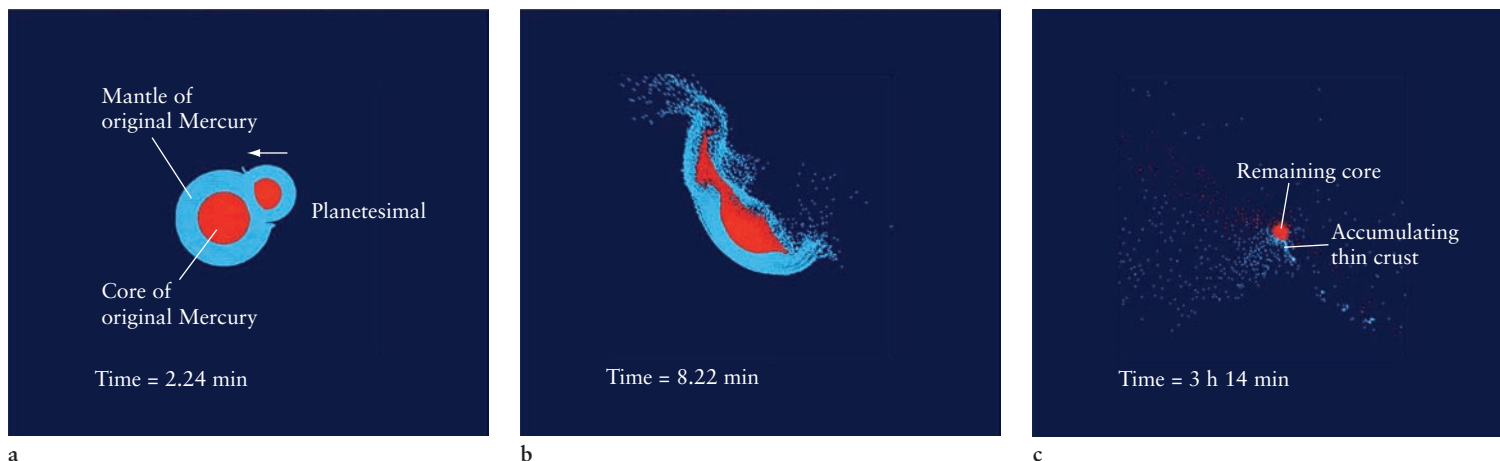


FIGURE 6-43 The Stripping of Mercury's Mantle To account for Mercury's high iron content, one theory proposes that a collision with a massive planetesimal stripped Mercury of most of its rocky mantle. These three images show a computer simulation of a nearly head-on collision between proto-Mercury and a body one-sixth its mass. Both

worlds are shattered by the impact, which vaporizes much of their rocky mantles. Mercury eventually reforms from the remaining iron-rich debris. The rest of the original Mercury and the impactor leave this area of the solar system. (Royal Astronomical Society)

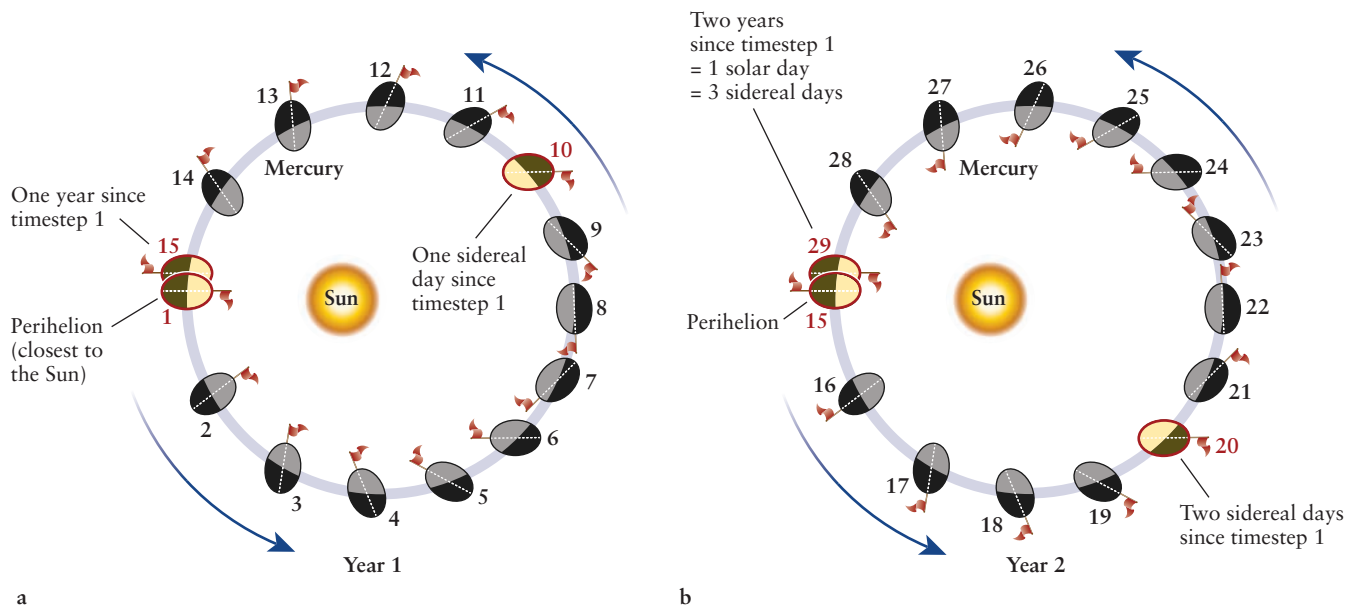


FIGURE 6-44 3-to-2 Spin-Orbit Coupling Mercury undergoes three sidereal rotations every 2 years. You can see this by following the flag on Mercury **(a)** from timestep 1 to timestep 15 and then **(b)** from timestep 15 to timestep 29. The flag points to the right four times during this interval (at times 1, 10, 20, and, finally, at 29). This means that Mercury has

rotated 3 times in exactly 2 sidereal Mercurian years. Because a sidereal year is 88 Earth days long, a sidereal day on Mercury is 58.7 Earth days long. During the same interval, however, the Sun is at noon as seen from the flag's location only twice: at timestep 1 and timestep 29. Therefore, a solar day is 176 Earth days long.

Insight Into Science

Model Building In modeling real situations, scientists must consider several different effects simultaneously. Omit any crucial property and you get inaccurate results. For example, consider how long ice can remain at Mercury's poles. Astronomers must take into account the planet's distance from the Sun, the tilt of its axis of rotation, its rotation rate, its surface features, its chemical composition, its atmosphere, and whether the ice on it is exposed or mixed with other material.

6-14 Mercury's atmosphere is the thinnest of all terrestrial planets

Mercury's mass is only 5.5% that of Earth. Like our Moon, the force of gravity is too weak on Mercury to hold a permanent atmosphere, but trace amounts of seven different gases—hydrogen, helium, sodium, calcium, magnesium, potassium, and oxygen—have been detected around Mercury. These gases are at least 10^{17} times less dense than the air we breathe. Scientists think the Sun is the source of the hydrogen and helium gas near Mercury, while sodium, magnesium, calcium, and potassium gases escape from rocks inside the planet (a process called *outgassing*, which also occurs on Earth). Different concentrations of these gases were

observed at different locations over the planet. Oxygen observed in Mercury's atmosphere may come from polar ice that is slowly sublimating. (Technically, when a solid, like ice, is transformed directly into a gas phase, the term is *sublimation*.) Impact of space debris also adds gases to the atmosphere. All of these gases drift into space and are continually replenished in the atmosphere from their respective sources. The radiation and particles flowing out of the Sun are pushing much of this gas outward, creating a tail for Mercury pointing away from the Sun. This tail is analogous to comet tails created by the Sun.

Focus Question 6-12

Why isn't Mercury in synchronous rotation with respect to the Sun?

5 Mercury's temperature range is the most extreme in the solar system Earth's thick atmosphere stores heat, which helps explain why it stays warm on cloudy days and at night. Nevertheless, Earth continually loses heat into space. Because of Earth's rapid daily rotation, this lost heat is quickly replaced by the Sun during daylight hours, so the average temperature change between day and night on Earth is only about 11 K (20°F). Mercury's slow rotation and its nearly complete lack of an atmosphere cause the difference in temperatures between day and night there to be far greater than on Earth. At noon on Mercury, the surface temperature is 700 K (800°F). At the terminator, where day meets night, the temperature is about 425 K (305°F). On the night side, shortly before sunrise, the temperature falls to as low as 100 K (−280°F), the coldest temperature on any terrestrial

planet! The resulting daily range of temperature on Mercury is therefore 600 K (1080°F).

VENUS

6 One of the few things that Mercury and Venus (Figure 6-45) have in common is that neither has a moon. Venus and Earth, on the other hand, have much in common. They have almost the same mass, the same diameter, and the same average density. Indeed, if Venus were located at the same distance from the Sun as is Earth, then it, too, might well have evolved life. However, Venus is 30% closer to the Sun than Earth, and this one difference between the two planets leads to a host of others, making Venus inhospitable to life.

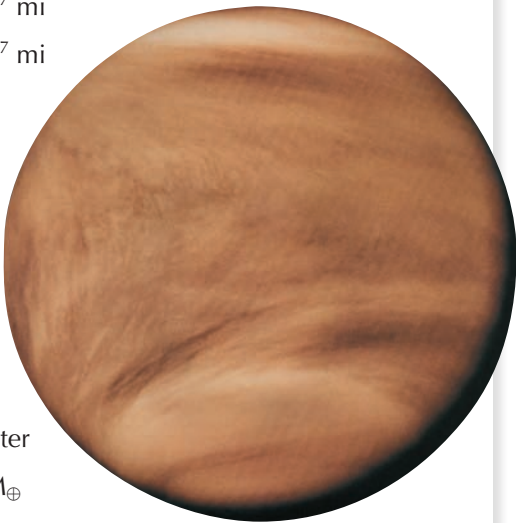
6-15 The surface of Venus is completely hidden beneath a permanent cloud cover

At nearly twice the distance from the Sun as Mercury, Venus is often easy to view without interference from the Sun’s glare. At its greatest western elongation, Venus is seen high above the western horizon after sunset, where, like Mercury, it is often called the “evening star.” Conversely, high in the eastern sky before sunrise, it is called the “morning star.” The maximum elongation of Venus is always between 45° and 47°.

Venus is easy to identify because it is usually one of the brightest objects in the night sky. Only the Sun and the Moon outshine Venus at its greatest brilliance. Venus is often mistaken for a UFO because when it is low on the horizon, its bright light is strongly refracted by

VENUS: Vital Statistics

Planet symbol:	♀
Average distance from the Sun:	0.723 AU = 1.082×10^8 km = 6.723×10^7 mi
Maximum distance from the Sun:	0.728 AU = 1.089×10^8 km = 6.767×10^7 mi
Minimum distance from the Sun:	0.718 AU = 1.075×10^8 km = 6.680×10^7 mi
Orbital eccentricity:	0.007
Average orbital speed:	35.0 km/s = 21.7 mi/s
Sidereal period of revolution (year):	224.7 days = 0.615 Earth year
Sidereal rotation period:	243.0 days (retrograde)
Solar rotation period (day):	116.8 Earth days
Inclination of equator to orbit:	177.4°
Inclination of orbit to ecliptic:	3.39°
Diameter (equatorial):	12,104 km = 7521 mi = 0.949 Earth diameter
Mass:	4.87×10^{24} kg = 1.07×10^{25} lb = $0.815 M_{\oplus}$
Average density:	5240 kg/m^3 = 8830 lb/yd ³ = 0.949 Earth density
Escape speed:	10.4 km/s = 6.5 mi/s
Surface gravity (Earth = 1):	0.91
Albedo:	0.75
Average surface temperature:	460°C = 860°F = 733 K
Surface air pressure:	92 × Earth’s surface air pressure
Atmospheric composition (by number of molecules):	96.5% carbon dioxide (CO ₂), 3.5% nitrogen (N ₂), 0.003% water vapor (H ₂ O)



R I V U X G

FIGURE 6-45 Venus’s Vital Statistics Venus’s thick cloud cover efficiently traps heat from the Sun, resulting in a surface temperature even hotter than that on Mercury. Unlike Earth’s clouds, which are made of

water droplets, Venus’s clouds are very dry and contain droplets of concentrated sulfuric acid. This ultraviolet image was taken by the Pioneer Venus Orbiter in 1979. (NASA)

Earth's atmosphere, making it appear to rapidly change color and position.

Unlike Mercury, Venus appears so bright in our sky because it is completely surrounded by light-colored, highly reflective clouds. Visible-light telescopes cannot penetrate this thick, unbroken layer of clouds, so we did not even know how fast Venus rotates until 1962. In the 1960s, however, both the United States and the former Soviet Union began sending probes there. The Americans sent fragile, lightweight spacecraft into orbit near the planet. The Soviets, who had more powerful rockets, sent more durable spacecraft directly into the Venusian atmosphere.

Building spacecraft whose technology could survive the descent into Venus's atmosphere proved to be more challenging than anyone had expected. Finally, in 1970, the Soviet probe *Venera* (Russian for "Venus") 7 managed to transmit data for 23 minutes directly from the Venusian surface. Soviet missions continued until 1985, measuring a blistering surface temperature of 750 K (890°F) and a surface air pressure of 92 atm, among other things. This value is the same pressure you would feel if you were swimming 0.82 km (2700 ft) underwater on Earth.

In contrast to Earth's present nitrogen- and oxygen-rich atmosphere (see Section 6-2), Venus's thick atmosphere is 96% carbon dioxide, with the remaining 4% mostly nitrogen. This atmosphere is remarkably similar to Earth's early carbon dioxide-rich atmosphere. These gases were vented from inside Venus through volcanoes and other openings. Unlike Earth, however, Venus has no liquid oceans to dissolve the carbon dioxide and no living organisms to convert it into oxygen and carbon compounds, so it remains in the atmosphere today.

Soviet spacecraft also discovered that Venus's clouds are confined to a layer that is 20 km (12 mi) thick and located 48 to 68 km (30 to 42 mi) above the planet's surface. Above and below the clouds are layers of haze that are also 20 km (12 mi) thick. Beneath the lower level of haze, the Venusian atmosphere is clear all the way down to the surface.

The results of the early probing of Venus's atmosphere are summarized in **Figure 6-46**. Both pressure and temperature decrease smoothly with increasing altitude. From the changes in temperature and pressure above a planet's surface, we can begin to understand the structure of its atmosphere.

Unlike the clouds on Earth, which appear white from above, the cloud tops of Venus appear yellowish or yellow-orange to the human eye. These colors are typical of sulfur and its compounds. Indeed, spacecraft found substantial amounts of sulfur dust in Venus's upper atmosphere and sulfur dioxide and hydrogen sulfide at lower elevations. The clouds themselves are composed

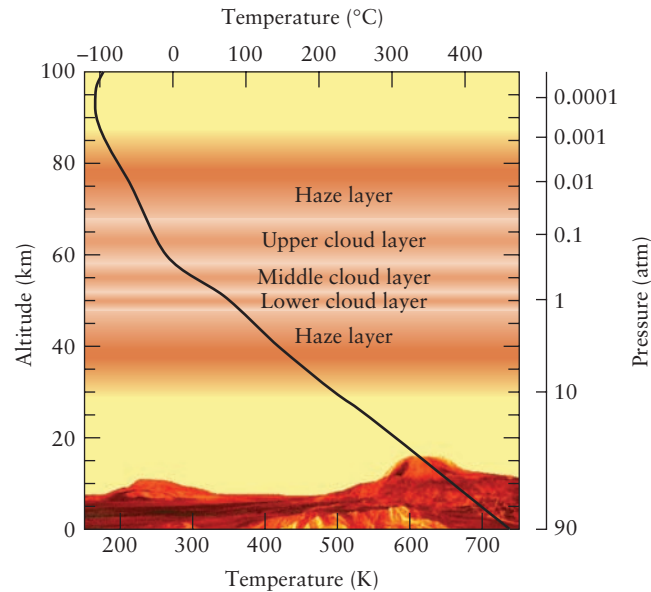


FIGURE 6-46 Temperature and Pressure in the Venusian Atmosphere The pressure at the Venusian surface is a crushing 92 atm (1296 lb/in²). Above the surface, atmospheric pressure and temperature decrease smoothly with increasing altitude. The temperature decreases from a maximum of nearly 750 K (890°F) on the ground to a minimum of about 173 K (−150°F) at an altitude of 100 km (62 mi). (NASA/Magellan Images [JPL])

of droplets of concentrated sulfuric acid! Because of the tremendous atmospheric pressure at Venus's surface, the droplets remain suspended as a thick mist, rather than falling as rain. Standing on the surface of Venus, you would experience a perpetually cloudy, yellow day (**Figure 6-47**).

All of the major chemical compounds spewed into our air by Earth's volcanoes have been detected in Venus's atmosphere. Among these molecules are large quantities of sulfur compounds. Because many of these substances are very short-lived, they must be constantly replenished by new eruptions. In fact, the abundance of sulfur compounds in Venus's atmosphere does vary. These data suggest the possibility that the sulfurous compounds in Venus's atmosphere come from active volcanoes. This idea is supported by the discovery, in 2007, of lightning on Venus. Lightning has many sources in atmospheres, including active volcanoes, as seen here on Earth when the Eyjafjallajökull (pronounced *ā-yuh-flah-lah-yolk*) volcano erupted in 2010 in Iceland (**Figure 6-48**). The lightning on Venus, caused by the electrical discharges in the atmosphere, created short-lived magnetic fields that were detected by the *Venus Express* satellite.

Focus Question 6-13

Why does the day have a yellow hue on Venus?



R I V U X G

FIGURE 6-47 The Venusian Surface The horizon and cloudy, carbon dioxide-rich sky of Venus imaged by the *Venera* lander. (Don P. Mitchell)



R I V U X G

FIGURE 6-48 Lightning on Earth This photograph of lightning generated by the Eyjafjallajökull volcano in Iceland in 2010 shows the phenomena that may have recently been observed on Venus. (David Jon/NordicPhotos/Getty Images)

6-16 The greenhouse effect heats Venus's surface

At first, no one could believe reports that the surface temperature on Venus was higher than the surface temperature on Mercury, which, after all, is closer to the Sun. Setting aside this initial skepticism, astronomers found a straightforward explanation—the **greenhouse effect** (Figure 6-49).

Carbon dioxide is responsible for the warming of Venus's atmosphere. Like your car windows, carbon dioxide is transparent to visible light but absorbs infrared radiation. Although most of the visible sunlight striking the Venusian cloud tops is reflected back into space, enough light reaches the Venusian surface to heat it. The warmed surface, which is also heated from inside the planet, in turn, emits infrared radiation, which is absorbed by Venus's carbon dioxide-rich atmosphere. Some of this trapped radiation returns to the planet's surface, producing the high temperatures found there. The rest of the heat is radiated into space.

7 Without the greenhouse effect, the surface of Venus would have a noontime temperature of 465 K (377° F). But with the greenhouse effect, that temperature is actually a sweltering 750 K (890° F), hotter than the hottest spot on Mercury! Indeed, Venus is the hottest of all four

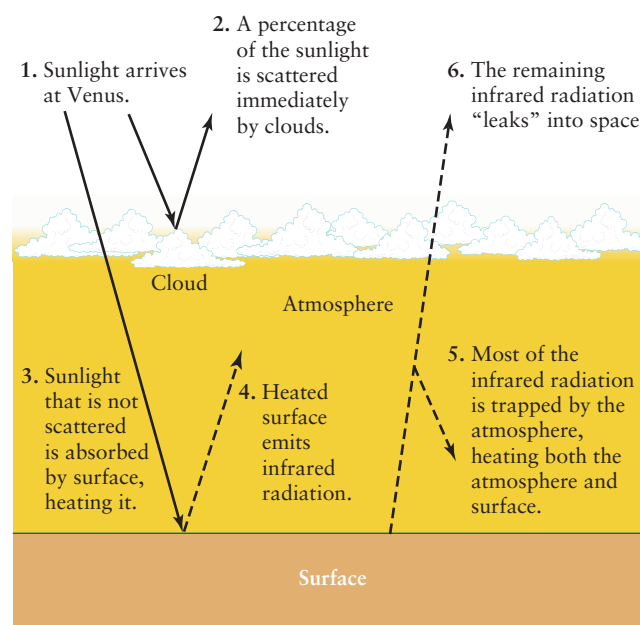


FIGURE 6-49 The Greenhouse Effect A portion of the sunlight penetrates through the clouds and atmosphere of Venus, heating its surface. The surface in turn emits infrared radiation, much of which is absorbed by carbon dioxide (and, to a much lesser degree, water vapor). The trapped radiation helps increase the average temperatures of the surface and atmosphere. Some infrared radiation does penetrate the atmosphere and leaks into space. In a state of equilibrium, the rate at which the planet loses energy to space in this way is equal to the rate at which it absorbs energy from the Sun.

terrestrial planets. Furthermore, the thick atmosphere keeps the night side of Venus at nearly the same temperature, unlike the night side of Mercury, where the temperature drops precipitously.

This high temperature prevents liquid water from existing on Venus. Just like Earth, Venus was likely to have initially had water both in the debris from which it formed and from the impacts of ice-bearing bodies, such as comets. The high atmospheric temperature has evaporated any water that came to Venus's surface, which is now bone dry. Today the total amount of water in Venus's atmosphere would uniformly cover the planet to a depth of only 3 cm (1 in). The equivalent depth of a uniform ocean on Earth today would be 2.8 km (1.7 mi). In 2007, using the *Venus Express* spacecraft, astronomers determined from the chemistry of its present atmosphere that Venus once had at least 150 times as much water. The atmospheric heat has caused most of the water that was once in its air to drift to the top of the atmosphere, where ultraviolet radiation from the Sun split the water molecules into hydrogen and oxygen. The light hydrogen and some of the oxygen drifted into space, while much of the oxygen combined with other elements in the atmosphere and on the surface of the planet.

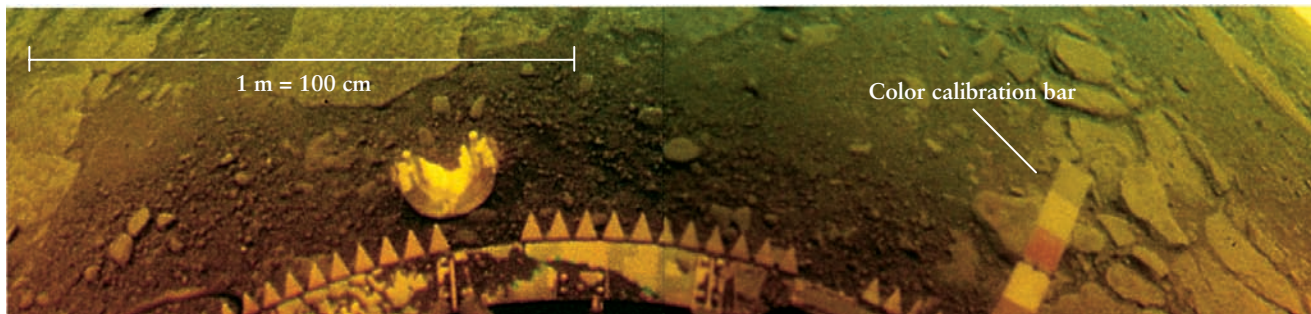
6-17 Venus is covered with gently rolling hills, two "continents," and numerous volcanoes

Soviet spacecraft that landed on Venus provided us with intriguing close-up images of the planet's arid surface. **Figure 6-50** is a view of Venus's regolith taken by *Venera 13* in 1982. Russian scientists believe that this region was covered with a thin layer of lava that contracted and fractured upon cooling to create the rounded, interlocking shapes seen in the photographs. Indeed, measurements by several spacecraft indicate that Venusian rock is quite similar to lava rocks, called *basalt*, which are common on Earth and the Moon.

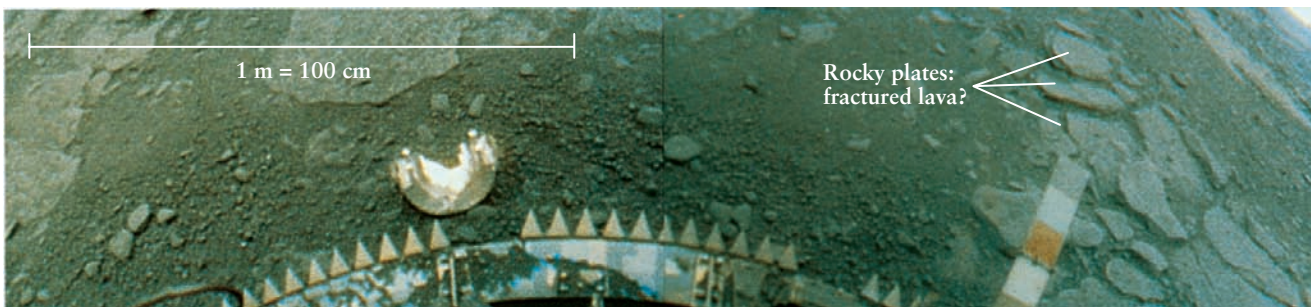
The entire Venusian surface was mapped by the *Magellan* spacecraft, which arrived at Venus in 1990. While in orbit about the planet, *Magellan* sent radar signals through the clouds surrounding Venus. Like radar used by police, some of the radar signals bounced off Venus and returned to the spacecraft. By measuring the time delay of the radar echo, scientists determined the heights and depths of Venus's hills and valleys. As a

Focus Question 6-14

Where on Earth (besides in a greenhouse) does the greenhouse effect occur?



a Image from *Venera 13*



b Color-corrected image

RTVUXIG

FIGURE 6-50 Color Correcting Light on Venus's Surface (a) This color photograph, taken by the Soviet *Venera 13* spacecraft, shows rocks that appear orange because the light was filtered through the thick, sulfur-rich clouds. (b) By comparing the apparent color of the spacecraft to the color it was known to be, computers can correct for the sulfurous light.

The actual color of the rocks is gray. In this view, the rocky plates that cover the ground may be fractured segments of a thin layer of lava. The toothed wheel in each image is part of the landing mechanism that keeps the spherical spacecraft from rolling. (Courtesy of C. M. Pieters and the USSR Academy of Sciences)

result, astronomers have been able to construct a three-dimensional map of the planet. The detail (or *resolution*) visible on this map is about 75 m (82 yd). You could see a football stadium on Venus if there were any. (None were detected.)

Venus is relatively flat compared to Earth. More than 80% of Venus's surface is covered with volcanic plains and gently rolling hills created by numerous lava flows. **Figure 6-51** is an image using the *Magellan* radar data to create a view of a Venusian landscape.

Global radar images revealed just two large highlands, or “continents,” rising well above the generally level surface of the planet (**Figure 6-52**). The continent at high latitudes in the northern hemisphere is called Ishtar Terra (after the Babylonian goddess of love) and is approximately the same size as Australia. Ishtar Terra is dominated by a high plateau, Lakshmi Planum, ringed by towering mountains. The highest mountain is Maxwell Montes, whose summit rises to an altitude of 11 km (6.8 mi) above the average surface. For comparison, Mount Everest on Earth rises 9 km (6 mi) above sea level.

The larger Venusian continent, Aphrodite Terra (named after the Greek counterpart of the Roman goddess Venus), is a belt of highlands that straddles the equator. Aphrodite Terra is 16,000 km (10,000 mi) in length and 2000 km (1200 mi) wide, giving it an area about one-half that of Africa. The global view of the surface of Venus in **Figure 6-53** shows that most of Aphrodite Terra is covered by vast networks of faults and fractures.

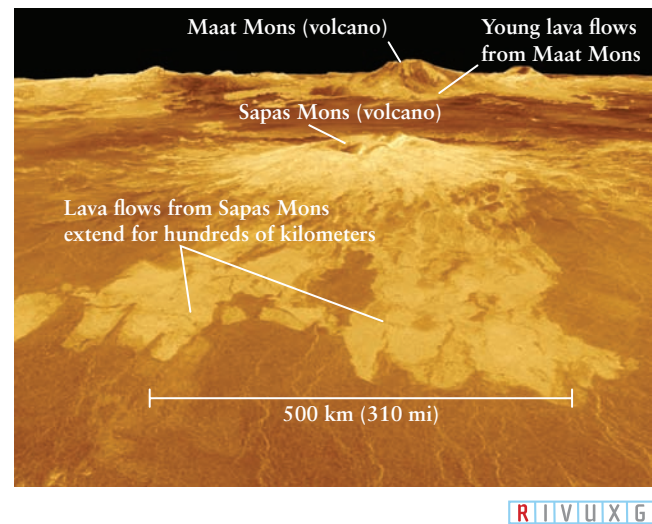


FIGURE 6-51 A Venusian Landscape A computer combined radio images to yield this perspective view of Venus as you would see it from an altitude of 4 km (2.5 mi). The color results from light being filtered through Venus's thick clouds. The brighter color of the extensive lava flows indicates that these flows reflect radio waves more strongly. The vertical scale has been exaggerated 10 times to show the gentle slopes of Sapas Mons and Maat Mons, volcanoes named for ancient Phoenician and Egyptian goddesses, respectively. (NASA, JPL Multimission Image Processing Laboratory)

Radar images reveal that Venus has more than 1600 major volcanoes and volcanic features and fewer than a thousand impact craters (**Figure 6-54**), as compared to the hundreds of thousands of impact craters seen on the

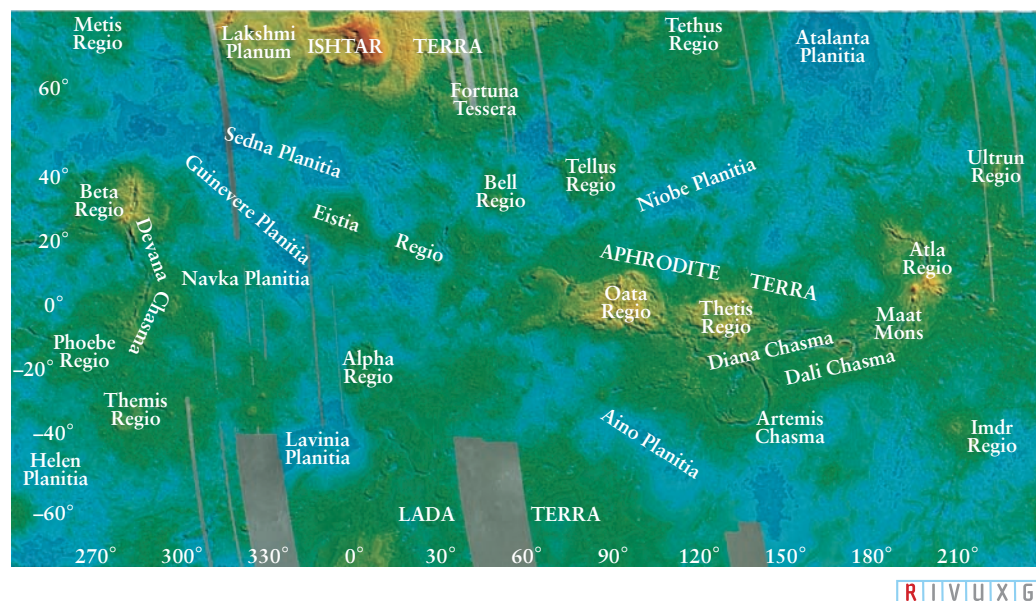


FIGURE 6-52 A Map of Venus This false-color radar map of Venus, analogous to a topographic map of Earth, shows the large-scale surface features of the planet. The equator extends horizontally across the middle of the map. Color indicates elevation—red for highest, followed by orange, yellow, green,

and blue for lowest. The planet's highest mountain is Maxwell Montes on Ishtar Terra. Scorpion-shaped Aphrodite Terra, a continentlike highland, contains several spectacular volcanoes. Do not confuse the blue and green for oceans and land. (Peter Ford/MIT, NASA/JPL)



R I V I U X I G

FIGURE 6-53 A “Global” View of Venus A computer using numerous *Magellan* images creates a simulated globe. Color is used to enhance small-scale structures. Extensive lava flows and lava plains cover about 80% of Venus’s relatively flat surface. The bright band running almost east-west is the continentlike highland region Aphrodite Terra. (NASA)

Moon and Mercury. Today, Venus’s thick atmosphere heats, and thereby vaporizes, much of the infalling debris that would otherwise create craters. Although pieces larger than around a kilometer in diameter would blast through even this thick atmosphere, most of the rubble entering Venus’s atmosphere is turned into dust by friction before it reaches the ground. However, there was most likely a period shortly after Venus formed and before its atmosphere developed (from gases escaping from the planet’s interior) when impacts and cratering were common.

If we accept that the young Venus was cratered like every other solid object in the solar system, we still need to explain the low number of craters on it today. Earth removes impact craters as a result of both weathering and recycling the crust via plate tectonic activity. Neither of these effects is in play on Venus, which has typical wind speeds of less than 7 km/h (5 mi/h), no rain, and no large-scale tectonic plates like those that divide up and refresh Earth’s surface. Put another way, Venus is a one-plate planet.

Without weather or tectonic plates moving relative to each other, how can the craters be erased? Several mechanisms to explain this are being explored. Most theories posit that the lack of surface motion has forced Venus’s crust to become thicker than Earth’s crust. In one theory, heat from radioactive elements inside the planet causes mantle convection without tectonic plate

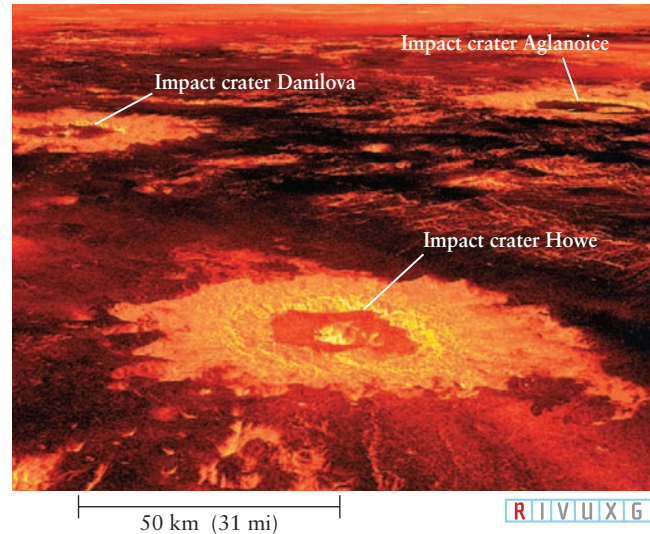


FIGURE 6-54 Craters on Venus These three impact craters, with extensive ejecta surrounding each, are located on Venus’s southern hemisphere. They were imaged using radar by the *Magellan* spacecraft. The colors are based on the *Venera* images (see Figure 6-50). (NASA/Magellan Images [JPL])

motion. The temperatures inside the planet eventually become high enough in the regions of rising mantle to create massive volcanoes out of which pour enormous amounts of magma that cover large areas of the surface. A variation of this theory proposes that the thickening crust is suddenly thinned from below by the horizontal convective motion of the mantle, like stripping layers off of plywood. When so thinned, heat is able to rise through the crust and cause many volcanoes to form.

Another theory proposes that, episodically, the one-plate surface of Venus is heated enough from below to crack and begin Earthlike tectonic plate motion until it cools down and again becomes a single plate. During the period of plate motion, the surface is erased and replaced.

In yet another theory, the mantle becomes so hot that large sections of the crust melt more or less simultaneously, destroying the old crust and its craters. Indeed, that theory has been expanded to periodic global meltdowns occurring every 700 million years or so. For all of these models, when enough heat is released from the planet’s interior, its surface once again solidifies and, with the possible exception of local volcanic activity, the crust rethickens.

The sulfur content of the air and the traces of active volcanoes on the surface are compelling indicators that, like Earth, Venus has a molten interior. Because the average density of Venus is similar to that of Earth, its core is predominantly iron. Currents in the molten iron should generate a magnetic field; however, other than the transient magnetic field detected by *Venus Express*, none of

Focus Question 6-15

Why does little cratering occur on Venus today, even compared to the present low rate of cratering on our Moon?

the spacecraft sent to Venus has detected one. In general, magnetic fields are created by rotation of currents (see Section 6-5). Therefore, Venus’s apparent lack of a magnetic field is plausible only if it rotates exceptionally slowly. In fact, the planet takes 116.8 Earth days to get from one sunrise to the next (if you could see the Sun from Venus’s surface).
That day is more unusual than just being much longer than ours. Unlike Earth and Mercury, Venus exhibits **retrograde rotation**, meaning that the direction of its rotation (clockwise as seen from space far above Earth’s North Pole) is *opposite* the direction of its orbit around the Sun, which is counterclockwise as seen from the same vantage point. As a result of its retrograde rotation, sunrise on Venus occurs in the west and sunset in the east.

Another perspective on its retrograde rotation is that Venus’s rotation axis is tilted more than 177°, compared to Earth’s 23½° tilt. That means that Venus’s north pole is on the other side of that world than Earth’s north pole. Because Venus’s axis is within 3° of being perpendicular to the plane of its orbit around the Sun, the planet has no seasons. Although we do not know the cause of Venus’s retrograde rotation, one likely explanation is that a monumental impact flipped the rotation axis early in the planet’s existence.

MARS

Mars’s distinctive rust-colored hue (Figure 6-55) makes it stand out in the night sky. For centuries Mars has

MARS: Vital Statistics

Planet symbol:	♂
Average distance from the Sun:	1.52 AU = 2.28×10^8 km = 1.42×10^8 mi
Maximum distance from the Sun:	1.67 AU = 2.49×10^8 km = 1.55×10^8 mi
Minimum distance from the Sun:	1.38 AU = 2.10×10^8 km = 1.30×10^8 mi
Orbital eccentricity:	0.093
Average orbital speed:	24.1 km/s = 15.0 mi/s
Sidereal period of revolution (year):	687 Earth days = 1.88 Earth years
Sidereal rotation period:	24 h 37 m 22 s
Solar rotation period (day):	24 h 39 m 35 s
Inclination of equator to orbit:	25.19°
Inclination of orbit to ecliptic:	1.85°
Diameter (equatorial):	6787 km = 4217 mi = 0.53 Earth diameter
Mass:	6.42×10^{23} kg = 14.2×10^{23} lb = $0.107 M_{\oplus}$
Average density:	3950 kg/m^3 = 6660 lb/yd^3 = 0.716 Earth density
Escape speed:	5.0 km/s = 3.1 mi/s
Surface gravity (Earth = 1):	0.38
Albedo:	0.15
Surface temperatures:	Maximum: 35°C = 95°F = 310 K Mean: −53°C = −63°F = 220 K
Surface air pressure:	Minimum: −140°C = −220°F = 133 K .006× Earth’s surface air pressure
Atmospheric composition (by number of molecules):	95.3% carbon dioxide (CO ₂) 2.7% nitrogen (N ₂) 0.03% water vapor (H ₂ O) 2% other gases

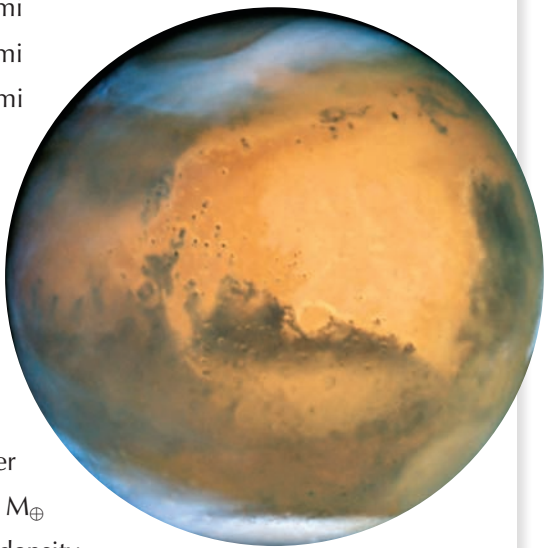


FIGURE 6-55 Mars’s Vital Statistics The lighter orange region in the center of the photograph in this figure is called Arabia Terra, after the Arabian peninsula on Earth. (NASA/USGS)

generated more excitement as a possible home for alien life than any other world in our solar system. The idea began following the first well-documented telescopic observations of Mars in 1659 by the Dutch physicist Christiaan Huygens. Huygens identified a prominent, dark surface feature that reemerged about every 24 hours, suggesting (correctly) a rate of rotation similar to that of Earth.

Belief that advanced life exists on Mars skyrocketed at the end of the nineteenth century after Giovanni Virginio Schiaparelli, an Italian astronomer, reported in 1877 seeing 40 lines crisscrossing the Martian surface (Figure 6-56). He called these dark features *canali*, an Italian term meaning “channels.” It was soon mistranslated into English as *canals*, implying the existence on Mars of intelligent creatures capable of engineering feats. This speculation led Percival Lowell, scion of a wealthy Boston family, to finance a major new observatory near Flagstaff, Arizona. By the end of the nineteenth century, Lowell had allegedly observed 160 Martian “canals.”

By the beginning of the twentieth century, it was fashionable to speculate that the Martian canals formed an enormous, planetwide irrigation network to transport water from melting polar ice caps to vegetation near the equator. In view of the planet’s reddish, desert-like appearance, Mars was thought to be a dying planet whose inhabitants must go to great lengths to irrigate

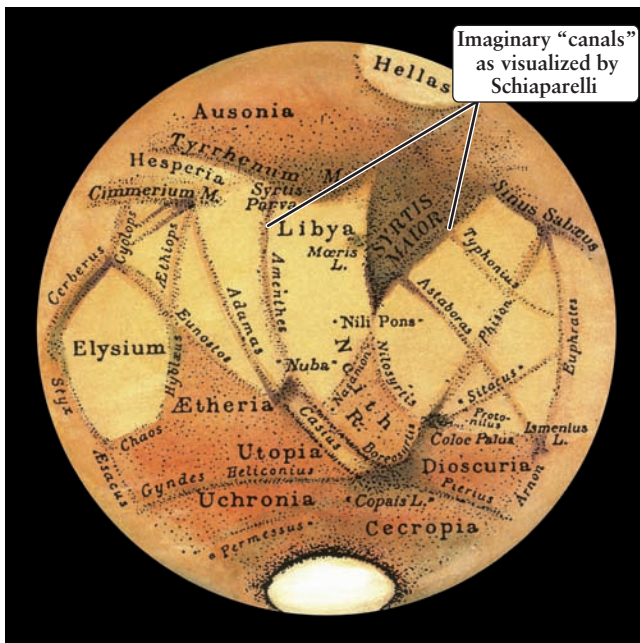


FIGURE 6-56 The Illusion of Martian Canals Giovanni Schiaparelli examined Mars through a telescope 20 cm (8 in) in diameter, the same size used by many amateur astronomers today. His drawings of the red planet showed features perceived by Percival Lowell and others as irrigation canals. Higher-resolution images from Earth and spacecraft visiting Mars failed to show the same features. (Detlev van Ravenswaay/Science Source)

Insight Into Science

Perception and Reality Our tendency to see patterns, even where they do not exist, makes it easy for us to leap to unjustified conclusions. Scientific images and photographs require objective analysis. This is why scientific discoveries are not considered to be authentic until they have been reproduced and verified by independent researchers.

their farmlands. No doubt the Martians would readily abandon their arid ancestral homeland and invade Earth for its abundant resources, an idea brought to life in 1898 by H. G. Wells in *War of the Worlds*. Ironically, we now have on the surface of Mars a spacecraft, *Curiosity*, which is zapping that planet’s surface with a laser reminiscent of the death rays depicted in the Wells story. That laser has zapped Mars’s surface over 100,000 times, vaporizing small amounts of it, and a spectroscope on *Curiosity* analyzes the chemical composition of the released gases. This analysis, combined with analysis of regolith scooped up by *Curiosity*, reveals the presence of perchlorate (a chlorine compound used in rocket fuel on Earth), water, a variety of rocks with unexpected chemical compositions, and relatively large grains of various minerals.

6-18 Mars’s global features include plains, canyons, craters, and volcanoes

As noted in Figure 6-55, Mars has only about 11% the mass and half the diameter of Earth. While Mercury’s small size and mass may have been due to a major collision when that planet was young, it appears that Mars is relatively small because Jupiter swept up much of the mass that would have gone into Mars, as discussed in Section 4-3.

Spacecraft journeying to Mars have sent back pictures since the 1970s. They show that Mars’s surface is completely dry and that it has broad plains, shallow craters, enormous inactive volcanoes, and vast canyons (Figure 6-57). The broad northern plain is called Vastitas Borealis (Latin for **northern waste** or **northern devastation**) and is shown in blue on Figure 6-57. This plain is some 5 km (3 mi) below the much more heavily cratered and hilly plains in the south, called the **southern highlands**. The cause of the significant difference in heights of the two hemispheres is not known, but it may stem from either early tectonic activity (see Section 6-3 for a discussion of tectonics) or from a powerful early impact.

Between the two hemispheres, *Mariner 9* discovered a vast canyon now called Valles Marineris, which runs roughly parallel to the Martian equator (Figure 6-58). Valles Marineris stretches over 4000 km (2500 mi), about one-fifth the circumference of Mars. If this canyon

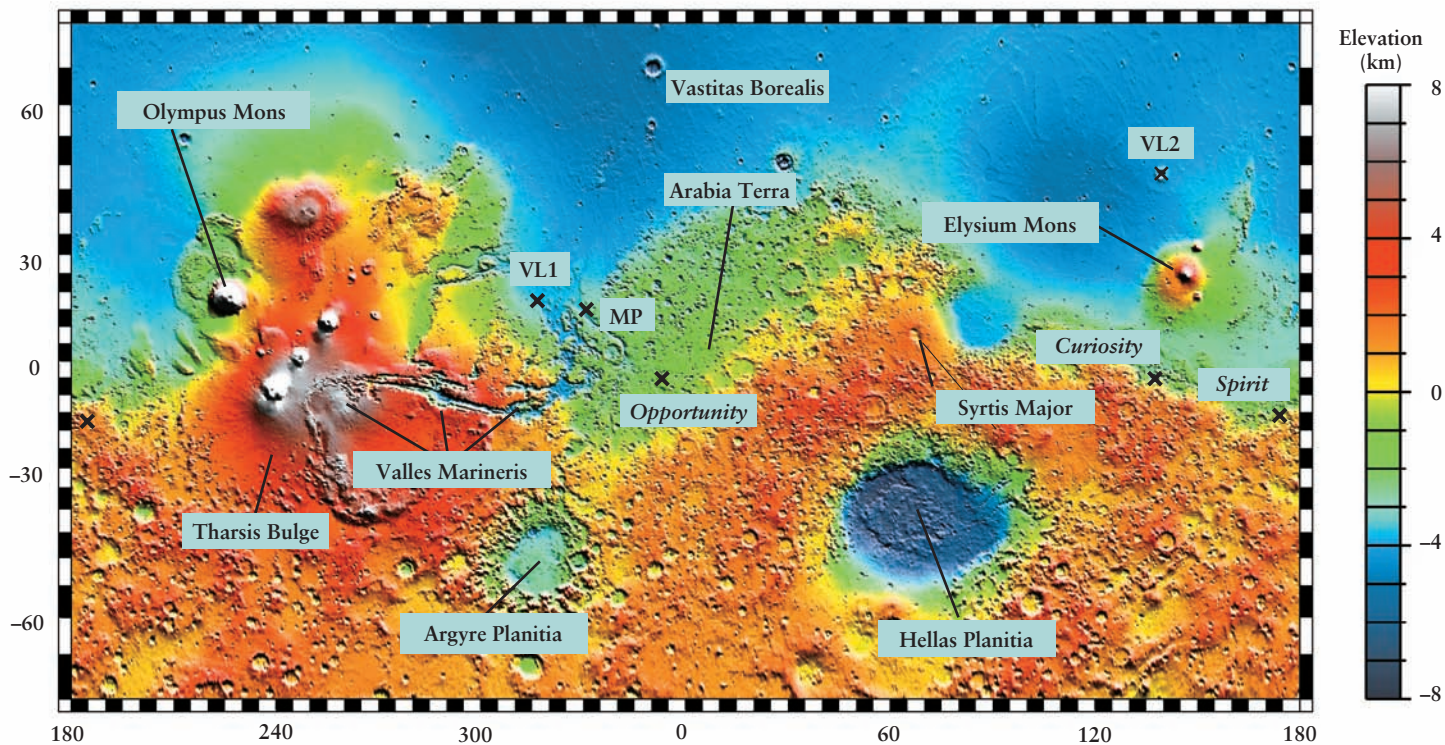


FIGURE 6-57 The Topography of Mars The color coding on this map of Mars shows elevations above (positive numbers) or below (negative numbers) the planet’s average radius. To produce this map, an instrument on board *Mars Global Surveyor* fired pulses of a light at the planet’s surface, then measured how

long it took each reflected pulse to return to the spacecraft. The *Viking 1 Lander* (VL1), *Viking 2 Lander* (VL2), *Mars Pathfinder* (MP), *Opportunity*, and *Spirit* landing sites are each marked with an X. The volcanoes depicted in white are higher than the scale on the right can show. (MOLA Science Team, GSFC/NASA)

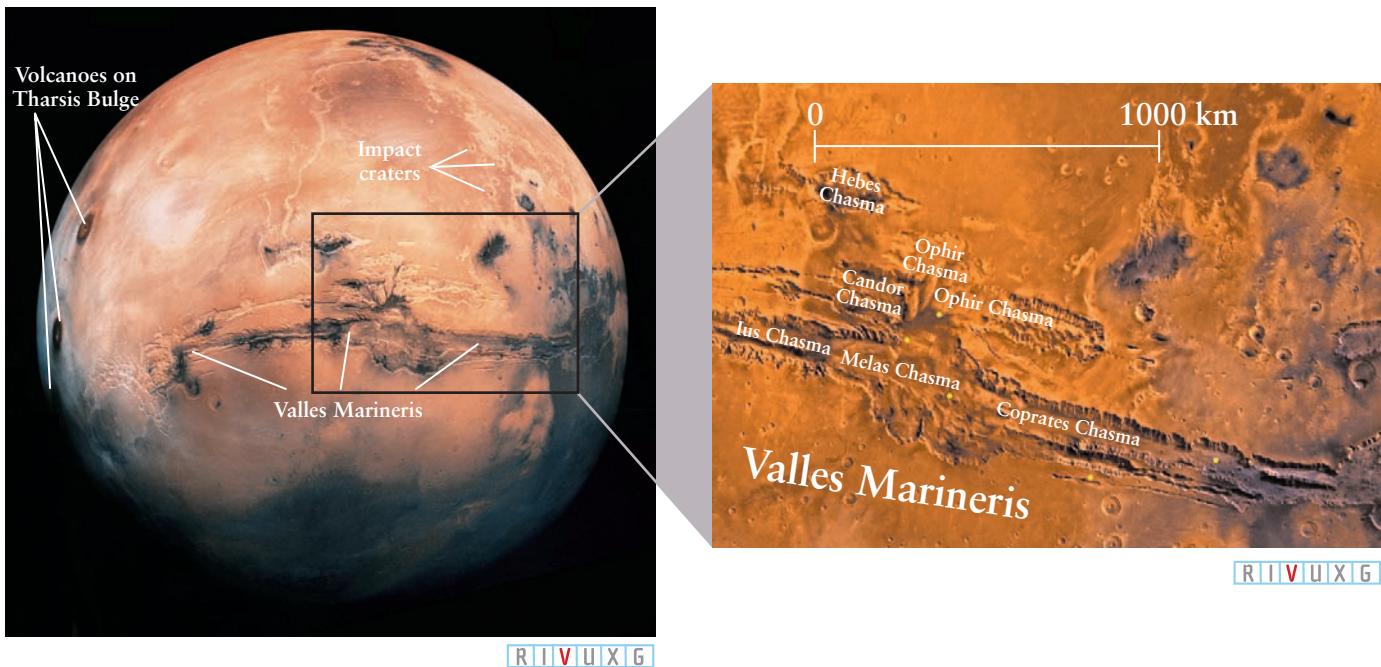


FIGURE 6-58 Martian Terrain This high-altitude photograph shows a variety of the features on Mars, including enormous volcanoes (left) on the highland, called Tharsis Bulge; impact craters (upper right); and vast, windswept plains. The enormous Valles Marineris canyon system crosses horizontally just below the center of the image. **Inset:** Details of the Valles

Marineris, which is about 200 km (120 mi) wide in this image. The canyon floor has two major levels. The northern (upper) canyon floor is 8 km (5 mi) beneath the surrounding plateau, whereas the southern canyon floor is only 5 km (3 mi) below the plateau. (USGS/NASA; inset: NASA/GSFP/LTP)

were located on Earth, it would stretch from New York to Los Angeles. The individual canyons that make up Valles Marineris are up to 8 km (5 mi) deep and 200 km (125 mi) wide. Valles Marineris begins with heavily fractured terrain in the west and ends with ancient-cratered terrain in the east. Geologists believe that the western part of Valles Marineris was created by tectonic plate motion on the young Mars. Data taken between 1999 and today reveal magnetic fields locked into parts of Mars's surface in just the same way that separating tectonic plates lock magnetic fields into Earth's rock on the boundaries between Earth's tectonic plates.

Some of the eastern parts of Valles Marineris appear to have been formed almost entirely by water flow, similar to Earth's Grand Canyon, which is 9 times shorter and 4 times less deep than Valles Marineris. Although remote observations such as those made by the *Mariner* and myriad other spacecraft have strongly suggested water has long existed on Mars, the chemical verification of water there, in the form of ice, was first made in 2008 by the *Phoenix Mars Lander*. In 2013, the *Curiosity* rover analyzed soil samples and determined that they contain 2% water.

Most impact craters are located in Mars's southern hemisphere (Figure 6-59). This suggests that the Vastitas Borealis has been resurfaced by some process that

eradicated ancient lowland craters, consistent with either of the earlier explanations of the differences in hemispheres.

Most of the volcanoes on Mars are located in the northern hemisphere. The largest volcano in the solar system, Olympus Mons (Figure 6-60a), covers an area as big as the state of Missouri and rises 26 km (16 mi) above the surrounding plains—nearly 3 times the height of Mount Everest. The highest volcano on Earth, Mauna Loa in the Hawaiian Islands, has a summit only 17 km (11 mi) above the ocean floor. Olympus Mons is a shield volcano, meaning that the lava poured down its sides, rather than being violently ejected. That manner of flow is evident in the inset of Figure 6-60a. While we do not know the active history of Olympus Mons, it is as high as it is in part because the lower gravity on Mars allowed the lava that issued from it to flow downhill more slowly than occurs on Earth. The lava solidified before it had spread out as much as occurs on Earth's shield volcanoes, thereby building up thicker layers and a taller volcano than the same flow would have created on Earth.

Around Olympus Mons and other volcanoes, orbiting spacecraft have observed numerous cratered, cone-shaped features that are typically the size of a

Focus Question 6-16

What two major activities created Valles Marineris?

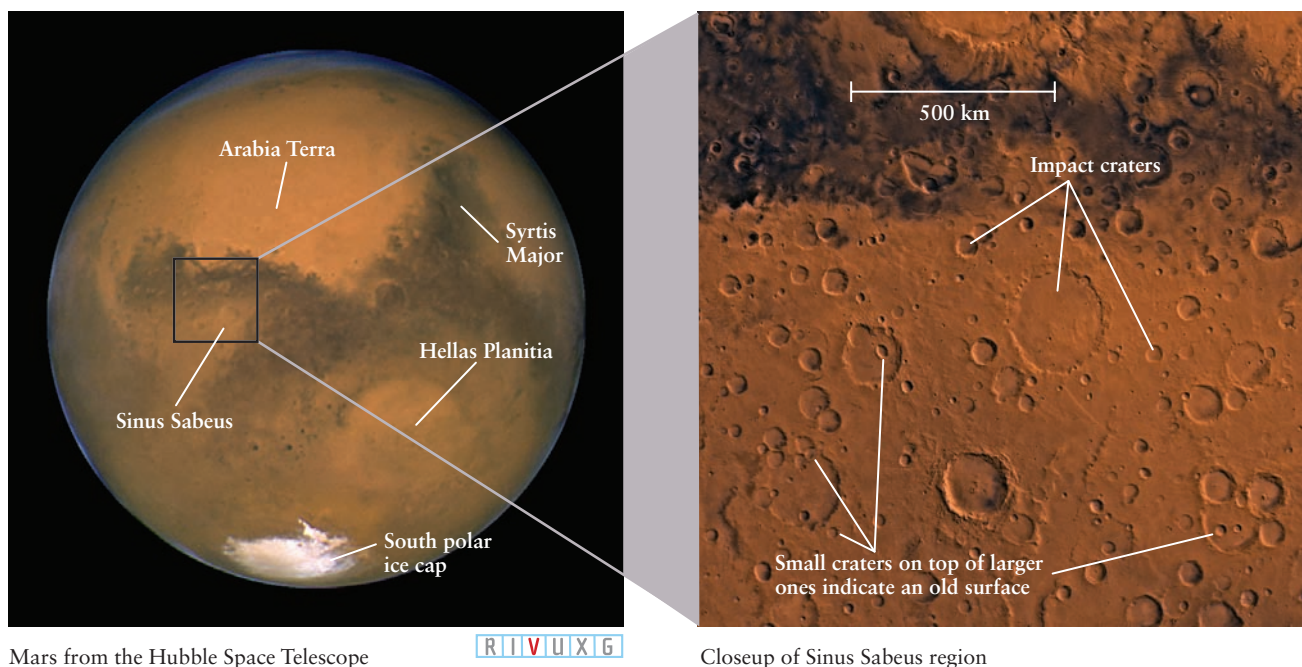


FIGURE 6-59 Craters on Mars This image taken by the Hubble Space Telescope was made during the opposition of 2003. Arabia Terra looks like the Arabian peninsula on Earth. It is an old, highly eroded region dotted with numerous flat-bottom craters. Lava-covered Syrtis Major was first identified by Christiaan Huygens in 1659. A single impact carved out

Hellas Planitia, which is 5 times the size of Texas. **Inset:** This mosaic of images from the *Viking 1* and 2 orbiter spacecraft shows an extensively cratered region located south of the Martian equator. Note how worn down these craters are compared to those on Mercury and our Moon. (NASA, J. Bell [Cornell U] and M. Wolff [SSI])

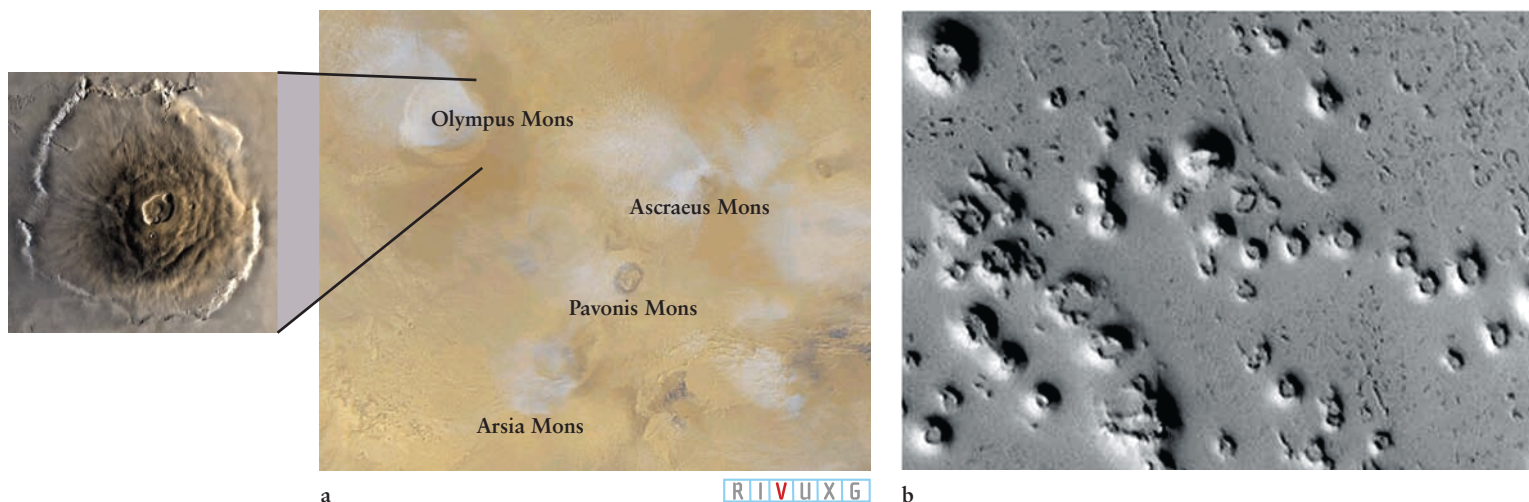


FIGURE 6-60 Volcanoes on Mars (a) Clouds of water vapor on Mars. **Inset:** Olympus Mons, the largest volcano on Mars (and in the solar system). (b) These cones on Mars may have been created when lava from Olympus Mons heated underground ice, causing the resulting

water and vapor to expand, raise the planet's surface, and burst out. (a: NASA; inset: NASA/JPL; b: Peter Lanagan [LPL, U. Arizona] et al., MOC, MGS, NASA)

football field (Figure 6-60b). These cones are believed to have been created by ice below the Martian surface that was liquefied and vaporized by lava from the nearby volcanoes. The expanding water pushed its way up to the surface, deforming the land before exploding out the centers of these “ice” volcanoes. Similar clusters of cones on Earth, formed by the same mechanism, are found in Iceland.

Astronomers have discovered impact craters on Mars that formed in the past few years. In one case, an image showed no craters, whereas an image taken 67 days later of the same area revealed a cluster of newly formed, small craters. On each California-sized patch of land on Mars, about two 1.0-m (1.1-yd) craters are created each year. Careful study of these new craters shows that they uncovered ice that had been buried only a meter or two under the planet's surface before the meteorite struck.

High-resolution photographs of Mars taken over the past 40 years, as shown on the previous pages, have failed to show one canal of a size consistent with those allegedly seen from Earth (see Figure 6-56). We now know for certain that Schiaparelli's *canali* were optical illusions, and the science fiction writers who wrote about advanced cultures there were completely off the mark.

Insight Into Science

Science Fact and Fiction Science fiction may foreshadow real scientific discoveries. (Can you think of some examples?) But science also reins in the fantasies of the science fiction writer. A century ago, millions of people, including top scientists, believed that technologically advanced life existed on Mars. However, all the scientific evidence says otherwise.

6-19 Although no canals exist on Mars, it does have some curious natural features

Mars not only lacks canals, but it also has no cities, roads, or other signs of civilization. However, astronomers have photographed several surface features that, at first glance, could have been crafted by intelligent life-forms. In 1976, a hundred years after Schiaparelli's alleged discovery of canals, the *Viking 1* orbiter spacecraft photographed a feature that appeared to be a humanlike face (Figure 6-61a). However, when the *Mars Global Surveyor* photographed the surface in greater detail in 1998 (Figure 6-61b), it found only natural structures: This feature was formed when softer rock eroded around a region of harder, less easily eroded rock.

Other photographs show a collection of pyramids, a skull (Figure 6-61c), and a “happy face” (Figure 6-61d). Are any of these features relics of an advanced civilization? Scientists universally believe that they are all naturally formed features. Like the face in Figures 6-61a and b, the pyramids and skull are consistent with winds eroding softer rock around harder rock pushed upward from inside Mars billions of years ago. The same thing happens on Earth. The happy face is made from a large impact crater and several small natural features inside it.

6-20 Mars's interior is less molten than the inside of Earth

Part of the Valles Marineris was apparently formed when large regions of Mars's surface moved apart due to tectonic plate motion. This activity is generated by the flow of rock under the planet's surface (see Section 6-3 for more on plate tectonics). However, there is no evidence that plate tectonic activity occurs on Mars today. This lack of activity indicates that more

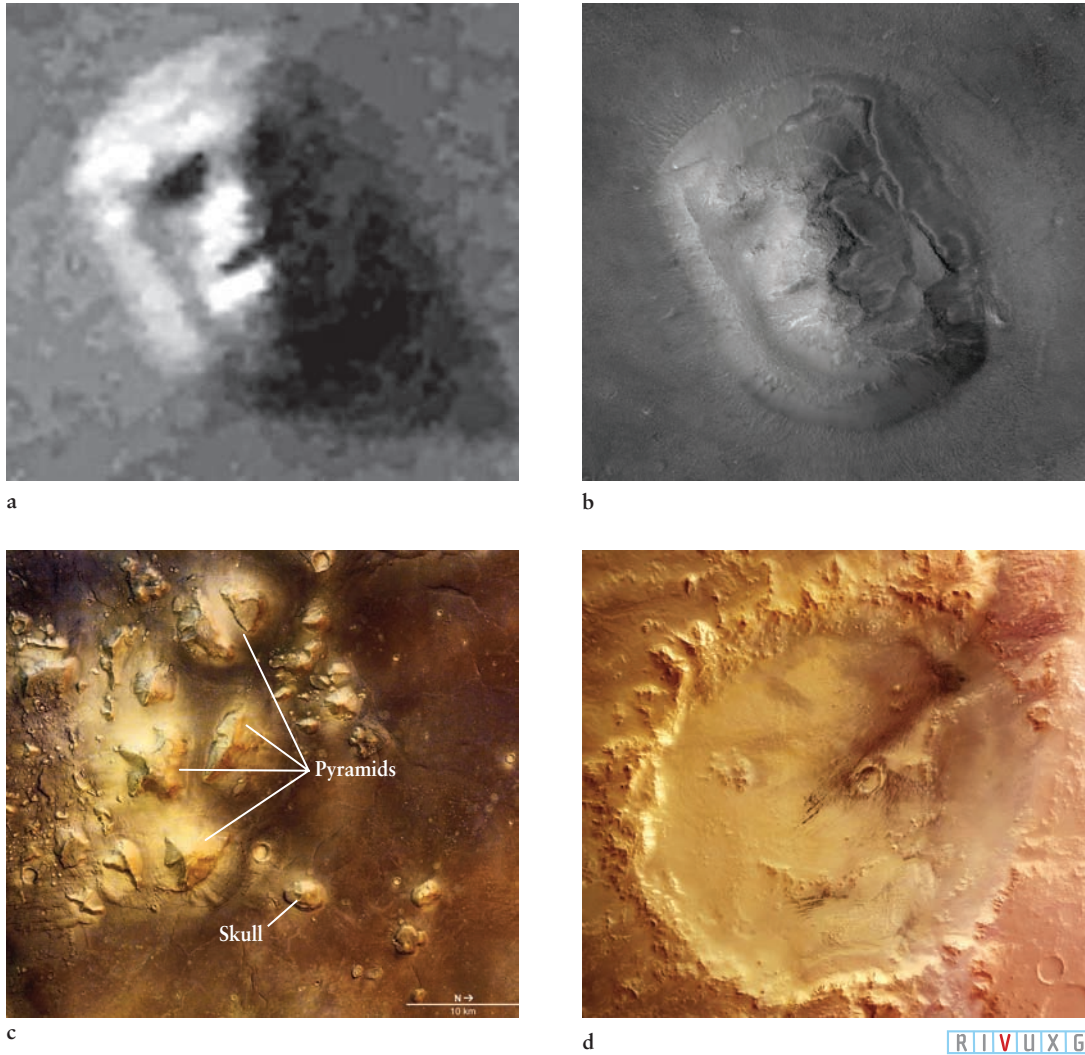


FIGURE 6-61 In the Eye of the Beholder Images (a) and (b) are of the same site on Mars, taken 22 years apart. They show how, at high resolution, the apparent face in (a) changed to a more “natural-looking” feature in (b). This change in appearance was due to improved camera technology. (c) In

the same region of Mars, other erosion features also appear to be pyramids and skulls. (d) The Galle crater and its interior features combine to give the impression of a “happy face.” (a: NSSDC/NASA; b, d: NASA/JPL/Malin Space Science Systems; c: European Space Agency/DLR/FU Berlin [G. Neukum]/Science Source)

of Mars’s interior is cooled solid rock than exists inside Earth. This conclusion is supported by the observation that Mars lacks a global magnetic field (see Section 6-5 for discussion of the field-generating dynamo effect). However, observations made in 2003 revealed that the Sun creates tiny tides (less than a centimeter) on the solid body of Mars. As the land rises and sinks ever so slightly through the day on Mars, friction created by this motion in the planet’s interior provides enough heat to keep some of the interior molten. (Rub your hands vigorously together to see how friction creates heat.)

Although a global magnetic field on Mars is absent today, there are remnants of such a field. In 1997 the *Mars Global Surveyor* (which operated until 2006) discovered patterns of local surface magnetic fields in

a variety of places on Mars (Figure 6-62), similar to regions on Earth where tectonic plates separate and lock Earth’s magnetic field in the rock (see Section 6-5). Geologists propose that when Mars was young, its internal “dynamo” created a strong magnetic field and that tectonic plates helped craft its surface while locking traces of its changing magnetic field in the then-molten rock. However, the planet quickly cooled, its global magnetic field vanished, and tectonic plate activity apparently ceased nearly 4 billion years ago.

A more molten interior early in its existence is also consistent with the enormous volcanoes that exist on Mars. Volcanoes are places where molten rock oozes or bursts from a body’s interior. Most of the volcanoes on Mars are

Focus Question 6-17

Have you ever seen something that you interpreted as one thing, but it turned out to be something else?

Insight Into Science

Keep It Simple II Was the Martian “face” made by intelligent life or by wind-sculpted hills? When two or more alternative interpretations present themselves, scientists apply the principle of Occam’s razor (see Section 2-1) and always choose the explanation that requires the fewest unproven assumptions.

the type where molten rock oozed out, called *shield volcanoes*, like the ones that created the Hawaiian Islands. Such volcanoes are very different from Mount Fuji in Japan or Mount Etna in Sicily, where the material from inside Earth is ejected violently. The first such volcano on Mars was discovered in 2013. On Earth, the Hawaiian Islands are only the most recent additions to a long chain of islands created by the growth of shield volcanoes rising through Earth’s crust under the Pacific Ocean. These volcanoes resulted from **hot-spot volcanism**, a process by which molten rock rises to the surface from a fixed hot region far below. The Pacific tectonic plate is slowly moving northwest at a rate of several centimeters per year. As a result, new volcanoes are created above the hot spot, while older ones move off and become extinct, eventually disappearing beneath the ocean. In the Atlantic Ocean, the Eyjafjallajökull volcano in Iceland (see Figure 6-48), which erupted in 2010, is located on a hot spot.

Because Mars’s surface is apparently frozen in place due to the lack of ongoing tectonic plate motion, a

hot spot can keep pumping lava upward through the same vent for millions of years. One result is Olympus Mons—a single giant volcano rather than a long chain of smaller ones (see Figure 6-60a). This volcano’s summit has collapsed to form a volcanic crater, called a **caldera**, large enough to contain the state of Rhode Island. Because Mars’s solid crust and mantle (the region directly below the surface) are now so thick, they insulate the surface, preventing any more molten rock from escaping. Therefore, we do not expect to see any more large-scale volcanic activity there.

The thickness of Mars’s solid crust and mantle and, hence, its ability to support great weight above it were emphasized in 2008, when radar on the *Mars Reconnaissance Orbiter* circling Mars observed that the planet’s surface does not buckle under the weight of its north polar ice cap. Earth has a thinner solid outer layer of crust and mantle, as indicated by the fact that our planet warps under similar ice layers. As a consequence of this discovery, astronomers have concluded that liquid water, if it exists under the northern Vastitas Borealis, is deeper inside the planet than previously thought. If that liquid water exists and there is life in it, then that life lies deep inside the planet.

The Martian crust, the outer part of its solid structure, is at least 40 km (25 mi) thick under the Vastitas Borealis and 70 km (43 mi) thick under the southern highlands. However, the boundary between the thin and thick crusts does not line up with the boundary between high and low terrain. Current models of Mars suggest that it has a core about 3400 km (2100 mi) in diameter

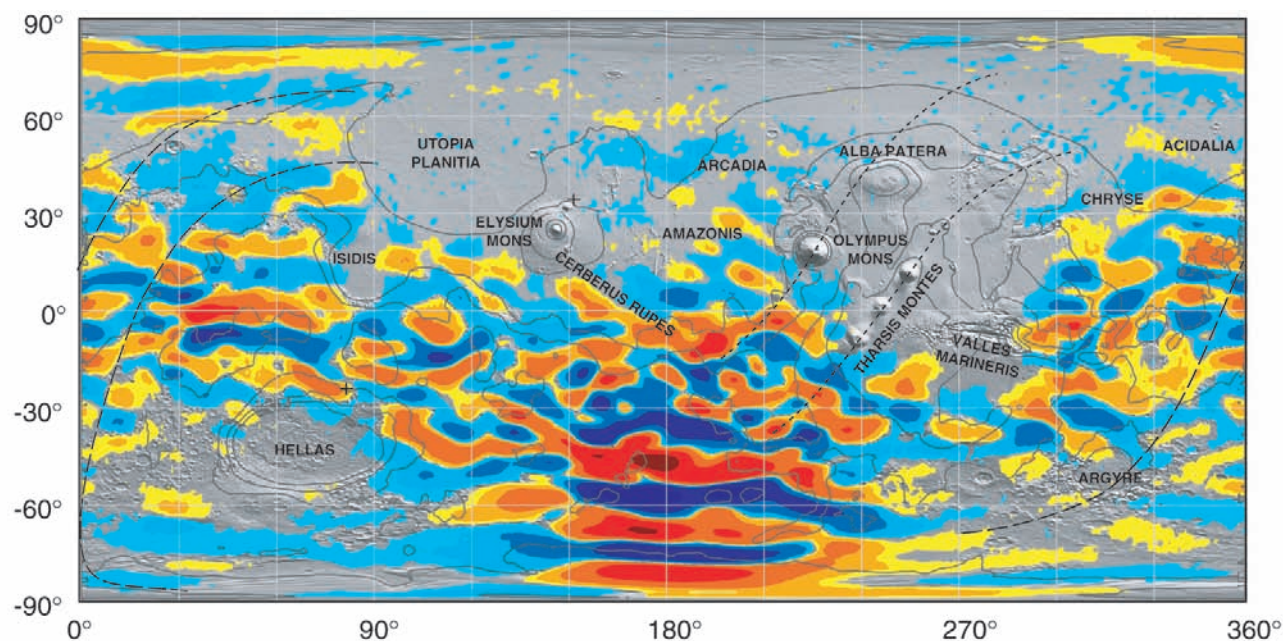


FIGURE 6-62 The Local Magnetic Fields on Mars Unlike Earth, with its global magnetic field, Mars has only local magnetic fields, shown here in the regions of yellow and

orange. They are believed to emanate from magnetized rock under the planet’s surface. (Jack Connerney / NASA)

and that at least some of that core is molten. A detailed understanding of Mars's interior waits for the placing of seismic detectors on the planet's surface.

6-21 Martian air is thin and often filled with dust

To understand more about the activity on Mars's surface, we must examine its atmosphere, which today has only 0.6% the pressure of Earth's atmosphere. As on Venus, some 95% of Mars's thin atmosphere is composed of carbon dioxide. The remaining 5% consists of nitrogen, argon, water vapor, and traces of oxygen. Unlike the gravitational attraction of tiny Mercury, which is too weak to hold any gases as a permanent atmosphere, Mars's gravitational force is just strong enough to prevent carbon dioxide, nitrogen, argon, and oxygen from escaping into space, but not strong enough to hold down water vapor.

A growing body of evidence indicates that there was once a vast water-filled ocean in the northern hemisphere of Mars and smaller bodies of water in numerous places over the entire planet. Evidence for the ocean includes layers of rock deposited by water and channels in which water from the south flowed northward. Observations made in 2008 suggest that liquid water existed on the surface of Mars for $2\frac{1}{2}$ billion years—in other words, until about 2 billion years ago. One of the factors that would have enabled the oceans to exist as long as they did was that Mars is believed to have had a much thicker atmosphere in the past. More air above it makes it harder for liquid water to evaporate. That atmosphere would have decreased in density for several reasons, including some gases in the upper atmosphere being heated enough to fly into space, some air molecules in the upper atmosphere being broken apart by ultraviolet radiation from the Sun and then drifting into space, some gases freezing and being buried under the surface, and some gases dissolving in the oceans.

Eventually, much of Mars's surface water evaporated and drifted into space, never to return. The rest either froze or was entrapped in underground lakes. Today, the concentration of water vapor in Mars's atmosphere is 30 times lower than it is in Earth's air. If all of the water vapor could somehow be squeezed out of the Martian atmosphere, it would not fill even one of the five Great Lakes of North America. The remainder of the water that was on Mars's surface has turned to ice, much of which remains there today. Most of that water has been covered with finely powdered, wind-blown regolith, which is why we do not see frozen lakes or similar features. Since 2002, satellites orbiting Mars have discovered enough water ice under Mars's southern pole to cover the entire planet to a depth of 11 m (36 ft). In 2008, the *Phoenix Mars Lander* dug into the Martian

regolith and observed ice just under the surface. Chemical tests made by it confirmed that the ice was frozen water, rather than frozen carbon dioxide or other compound. Water ice is being discovered at many latitudes from the polar regions to the equator. In 2008, the *Mars Reconnaissance Orbiter* discovered buried glaciers extending dozens of kilometers from mountains and cliffs. It is certain that much more water remains to be found inside Mars.

Despite the low density of Mars's air, the red planet experiences Earthlike seasons because of a striking coincidence, first noted in the late 1700s by the German-born English astronomer William Herschel. Just as Earth's equatorial plane is tilted $23\frac{1}{2}^\circ$ from the plane of its orbit, Mars's equator makes an angle of about 25° from the plane of its orbit. If, as is now suspected, the northern hemisphere of Mars is lower than the southern hemisphere because the north was struck by a large body, that impact could have also tilted the planet and initiated the seasons there. The Martian seasons last nearly twice as long as Earth's, because Mars takes nearly 2 Earth years to orbit the Sun. When Mars is near opposition, even current telescopes for home use reveal its daily changes. Dark seasonal markings on the Martian surface can be seen to vary, and prominent polar caps of frozen carbon dioxide shrink noticeably during the spring and summer months (Figure 6-63).

In 1994, astronomers using computer simulations discovered that the axis tilts of rapidly rotating planets, like Mars and Earth, are prone to change over millions of years unless such a planet is stabilized by the presence of a massive moon. Earth has such a moon; hence, our planet maintains its present $23\frac{1}{2}^\circ$ tilt, but the two tiny moons of Mars do not have enough gravitational attraction to keep the orbital inclination of Mars from changing dramatically. Observations of the underground polar regions by the *Mars Reconnaissance Orbiter* reveal seasonal changes on Mars and changes in the chemistry of its atmosphere over periods of millions of years consistent with changes in the tilt of Mars's rotation axis far from its present 25° .

Although sometimes blue (Figure 6-64a), Mars's atmosphere is often pastel red, sometimes turning shades of pink and russet (Figure 6-64b). All of these latter colors are due to fine dust blown from the planet's desertlike surface during windstorms. The dust is iron oxide, familiar here on Earth as rust. Mars's sky also changes color because the amount of dust in the air varies with the season. During the winter, carbon dioxide ice adheres to the dust particles and drags them to the ground. This helps clear and lighten the air. In the summer months, the carbon dioxide is not frozen, and the dust blown by surface winds remains aloft longer.

Focus Question 6-18

If the Hawaiian Islands were fixed over their hot spot, rather than moving, what would happen to them?

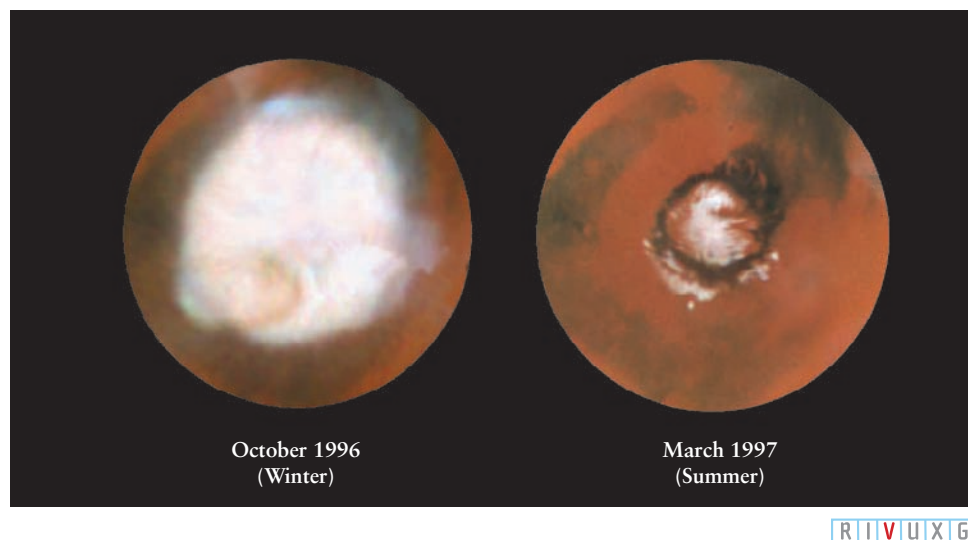


FIGURE 6-63 Changing Seasons on Mars During the Martian winter, the temperature decreases so much that carbon dioxide freezes out of the Martian atmosphere. A thin coating of carbon dioxide frost covers a broad region around Mars's north pole.

During the summer in the northern hemisphere, the range of this north polar carbon dioxide cap decreases dramatically. During the summer, a ring of dark sand dunes is exposed around Mars's north pole. (S. Lee/J. Bell/M. Wolff/Space Science Institute/NASA)

The sky color also changes over periods of many years. In 1995, for example, the amount of dust was observed to have dropped dramatically compared to that observed in the 1970s. The reason for such long-term changes is still under investigation.

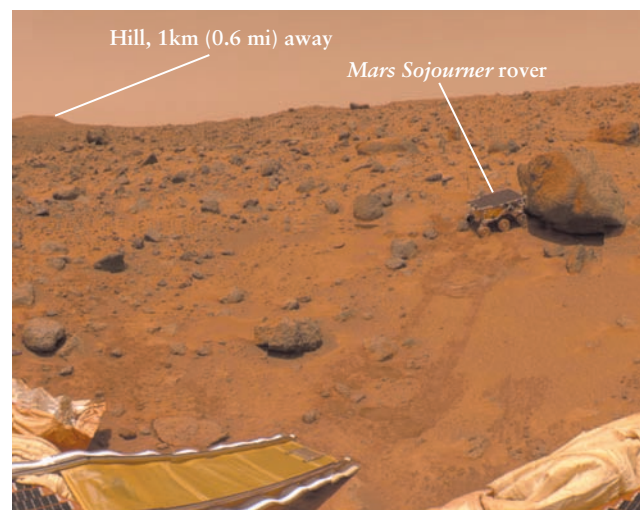
As on Earth, the temperature changes with the seasons on Mars. For example, one day at the landing site of the rover *Spirit*, near the equator, the tempera-

ture ranged from 188 K (-120°F) to 243 K (-22°F), whereas 100 days later it ranged from 200 K to 263 K (-100°F to 14°F , see Figure 6-57 for the site location). During the afternoons, heat from the planet's surface warms the air and sometimes creates whirlwinds, called **dust devils** (Figure 6-65). A similar phenomenon occurs in dry or desert terrain on Earth. Martian dust devils reach altitudes of 12 km (7.5 mi). Spacecraft on Mars



a

FIGURE 6-64 The Atmosphere of Mars (a) When Mars's sky is relatively free of dust, it appears similar in color to our sky, as shown in this sunset photo taken by the rover *Opportunity*. The darker, brown color in which the Sun is immersed is due to lingering dust in the sky. When less dust is present, the Sun looks almost white during Martian sunsets. Most of the

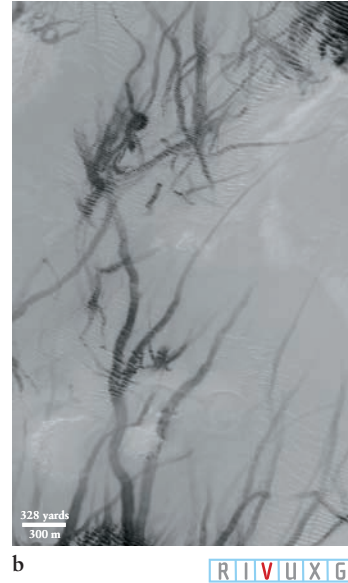


b

images we have of Mars's sky show colors like those seen in (b). (b) Taken by the *Mars Pathfinder*, this photograph shows the *Sojourner* rover snuggled against a rock named Moe on the Ares Vallis to run tests on the rock. At the top of the image, the pink color of the Martian sky is evident. (a: NASA/JPL; b: NASA/Lunar and Planetary Institute)



FIGURE 6-65 Martian Dust Devil (a) This *Spirit* image is one in a movie sequence of a dust devil moving left to right across the surface of Mars. The rovers have filmed several such events. (b) These dark streaks are the paths of dust devils on the Argyre Planitia of Mars. The tracks cross hills, sand dunes, and boulder fields, among other features on the planet's surface. (a: NASA/JPL; b: NASA/JPL/Malin Space Science Systems)



detect drops in air pressure as dust devils sweep past, just as on Earth. Some dust devils are large enough to be seen by orbiting spacecraft.

The winds on Mars help explain why its surface is eroding. Over the past two centuries, astronomers have seen faint surface markings disappear under a reddish-orange haze as thin Martian winds have stirred up finely powdered dust from the Martian regolith. Some storms obscure the entire planet, as happened in 2001. Over the ages, winds have worn down crater walls, and deposits of dust have filled in the crater bottoms (Figure 6-66). But the Martian atmosphere is so thin that 3 billion years of sporadic storms have not carried enough of Mars's extremely fine-grained powder to eradicate the craters completely.

6-22 Surface and underground features indicate that water once flowed on Mars

Despite disproving the theory that Mars has broad canals or any other liquid surface water, Mars-orbiting space-

craft did reveal many dried-up riverbeds (Figure 6-67a); river deltas (Figure 6-68), where water and debris emptied into lakes and oceans; lakes (Figure 6-68a); sedimentation laid down by water flow (Figure 6-69); and other water-related features. Some of the riverbeds include intricate branched patterns and delicate channels meandering among flat-bottomed craters. Rivers on Earth invariably follow similarly winding courses (Figure 6-67b).

Surface rovers *Spirit* and *Opportunity* found strong evidence supporting the belief that water created many of these features (Figure 6-70), including the presence of water-formed hematite—iron-rich rocks that look like blueberries—and rocks with extremely high levels of salts that had previously been dissolved in liquid water. Such salts take millions of years to seep into rocks in the quantities they have been found, implying that liquid water was on the red planet's surface for at least that length of time. The rover *Curiosity* rode over a riverbed

Focus Question 6-19

Why do you think it is easier to stand up in a dust devil on Mars than in the same equivalent-speed event on Earth?

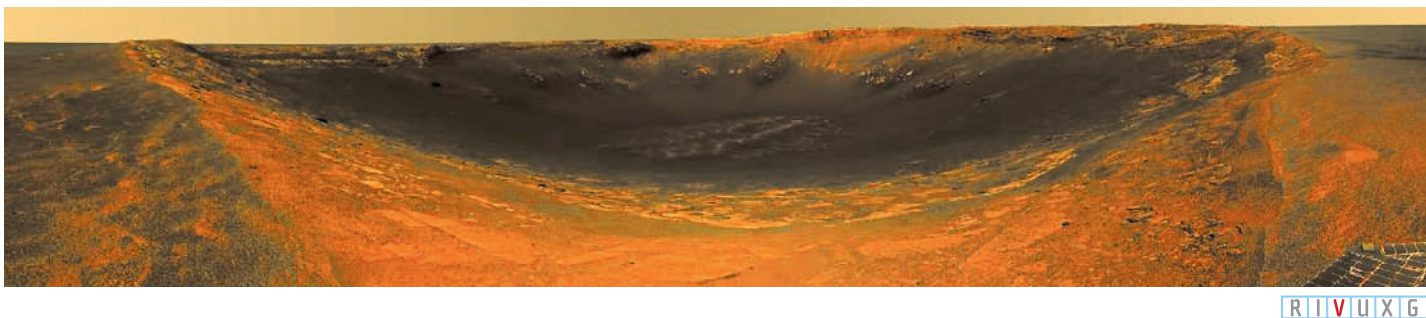


FIGURE 6-66 Crater Endurance Photographed by the rover *Opportunity*, this crater on Mars is about 130 m (430 ft) across. Rocks are visible in the crater and in vertical cliffs

along its walls. By studying such rocks, astronomers hope to understand more of the history of water on Mars. (JPL/NASA)

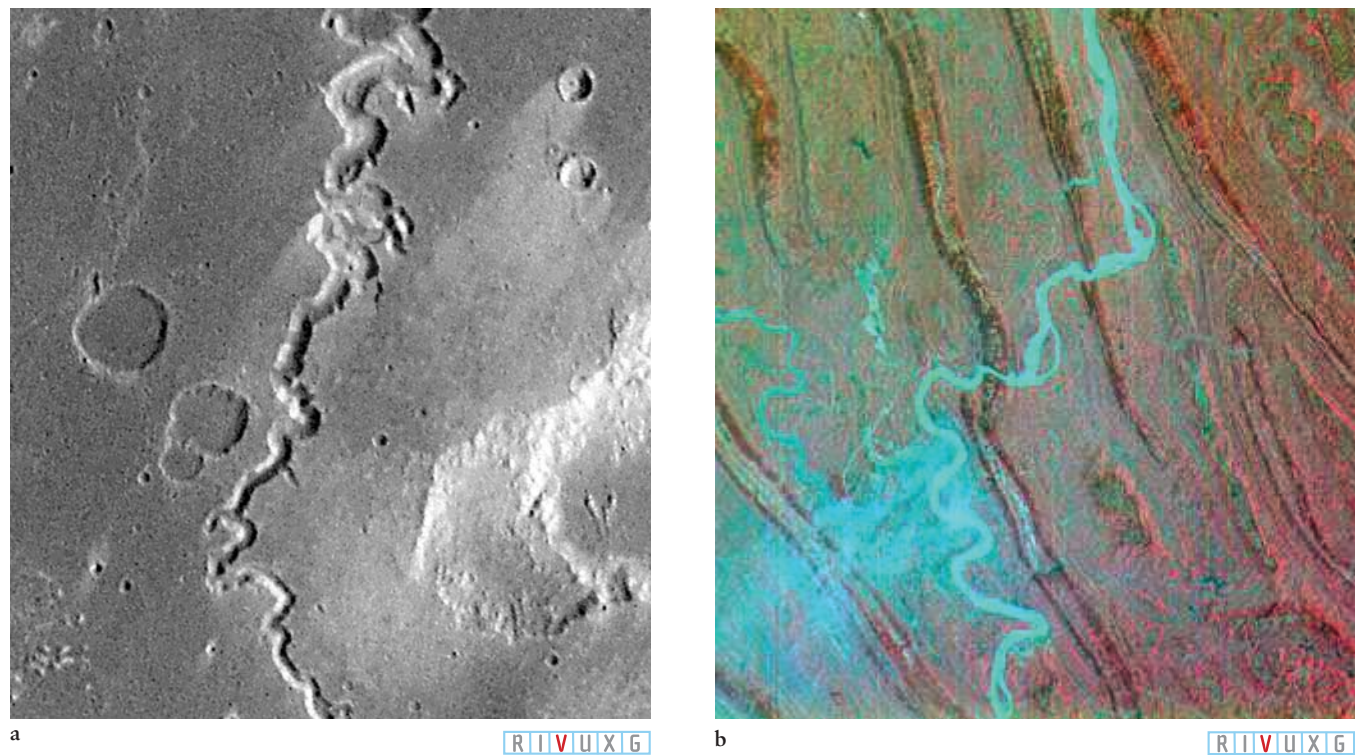


FIGURE 6-67 Rivers on Mars and Earth (a) Winding canyons on Mars, such as the one in this *Viking 1* orbiter image, appear to have been formed by sustained water flow. This theory is supported by the terraces seen on the canyon walls in high-resolution *Mars Global Surveyor* images. Long

periods of water flow require that the planet's atmosphere was once thicker and its climate more Earthlike. (b) The Yangtze River near Chongqing, China. Typical of rivers on Earth, it shows the same snakelike curve as the river channels on Mars. (a: NASA; b: TMSC/NASA)

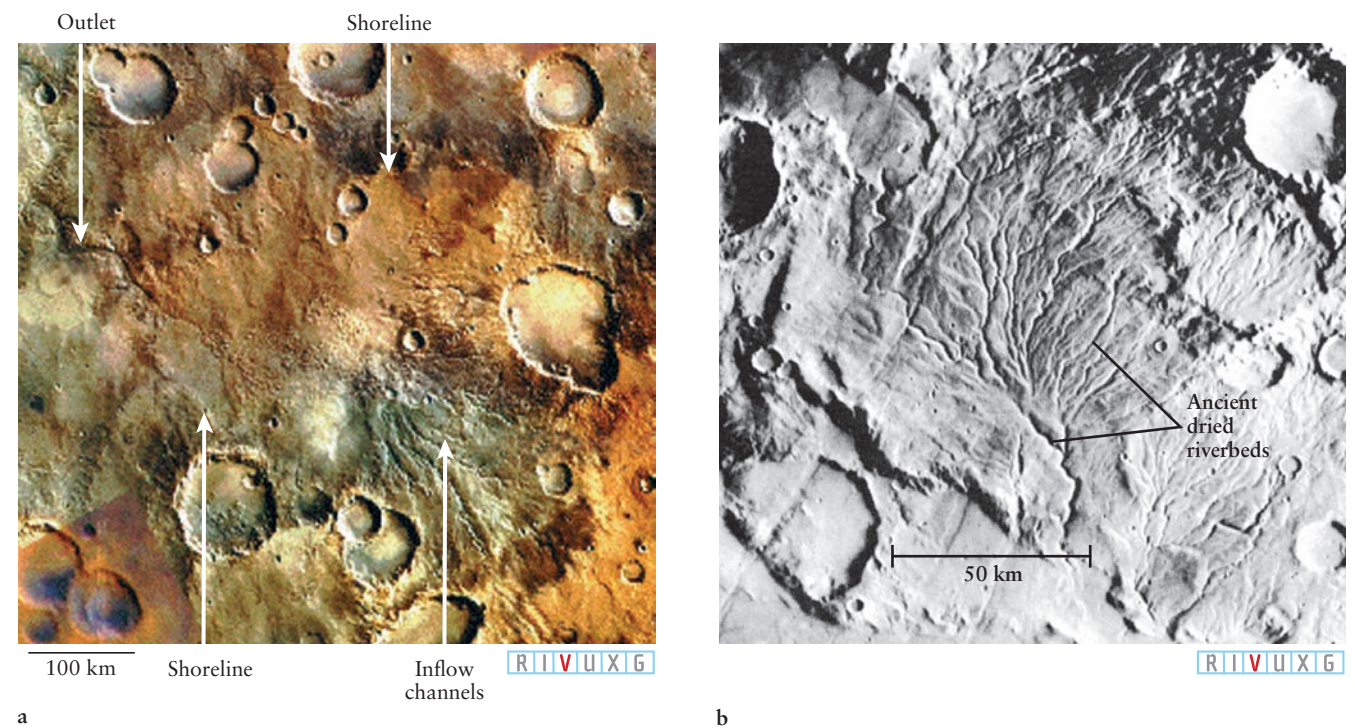


FIGURE 6-68 Evidence of Water on Mars (a) This dry Martian lake, photographed by the *Mars Global Surveyor* with a resolution of 1.5 m (1.6 yd), is an excellent example of how geology and astronomy overlap. The

features of this dry lake are consistent with those found on lakebeds on Earth. (b) This network of dry riverbeds is located on Mars's cratered southern hemisphere. (a: NASA/JPL; b: NASA)

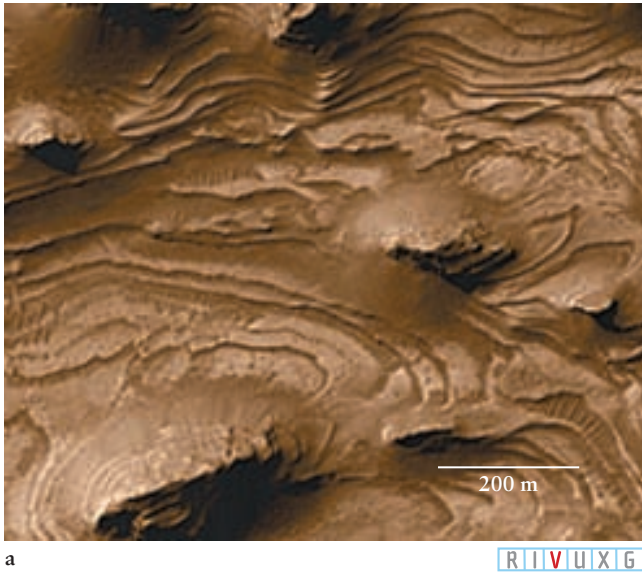
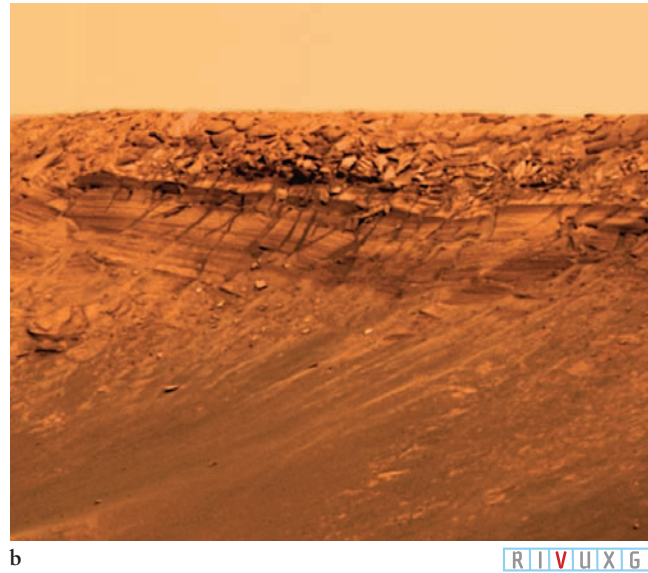


FIGURE 6-69 Ancient Oceans and Lakes on Mars (a) This *Mars Global Surveyor* image of a portion of Valles Marineris reveals terrain with “stair-step layers.” Such terrain is likely to have been created by sedimentation at the bottom of an ancient



body of water. (b) This image of Burns Cliff, photographed by the rover *Opportunity*, shows a close-up of layers of rock laid down on a body of water on Mars that went through wet and dry periods. (a: NASA/JPL/Malin Space Science Systems; b: NASA/JPL/Cornell)

in 2012, adding more credence to the belief that water once flowed on the red planet. On the other hand, observations show that the clay deposits laid down by liquid water on some parts of Mars are relatively thin, suggesting that while water was on the surface in some places for billions of years, there are other places on the surface where it lasted for much shorter lengths of time. Some of the clay on Mars appears to have come from lava flows,

weakening further the case for extended liquid water coverage everywhere on the planet. About 6% of Mars’s surface is covered with material of volcanic origin.

The origin of at least some of the water that was on Mars is believed to have been ice-rich bodies that struck the surface, releasing their water into the atmosphere, where it later rained down on the surface. Although water existed on Mars’s surface for long periods, there

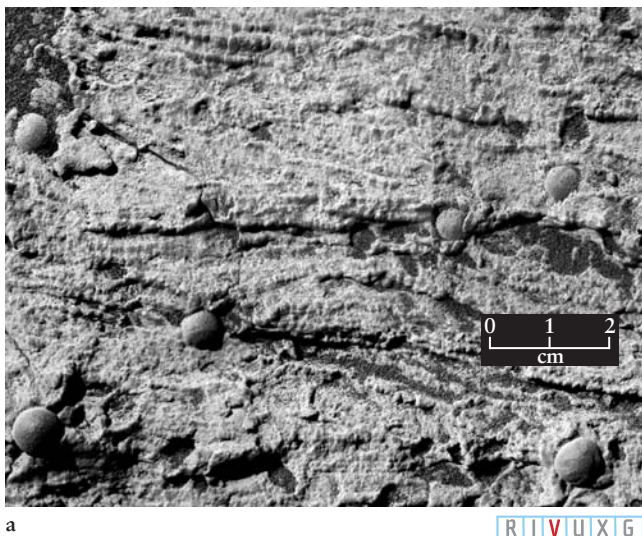


FIGURE 6-70 Layers of Rock Laid Down by Water (a) This close-up image taken by the rover *Opportunity* shows a small section of rock layers in a location called the Dells. The angled and curved layering seen here is created only on Earth by water flow, strongly suggesting that this sediment was also deposited by water. The nearly spherical rocks, called “blueberries” because they are dark, have



been chemically identified as hematite, an iron-rich mineral that is usually formed in water. The rovers have found blueberries strewn in a wide variety of locations. (b) This iron meteorite, dubbed the “Heat Shield Rock” because it was discovered behind the Mars rover *Sojourner*’s heat shield, is surrounded by hematite blueberries. (a: NASA/JPL/USGS; b: NASA/JPL/Cornell)

are indications in several places that large bodies of water have been gone for billions of years, as mentioned earlier. The evidence for this is the discovery by the rovers of lava rocks that contain olivine and pyroxene, rocks that dissolve quickly in water. If these rocks came to the surface while there was still water on it, these two minerals in the rock would have been dissolved away.

The surface of Mars continues to change. For instance, images taken by *Mars Global Surveyor* of one crater wall show that, sometime between 1999 and 2005, there was a significant landslide on it (**Figure 6-71a, b**). While this new feature was initially attributed to liquid water gushing out and initiating the landslide, subsequent observations of similar activity suggest that the

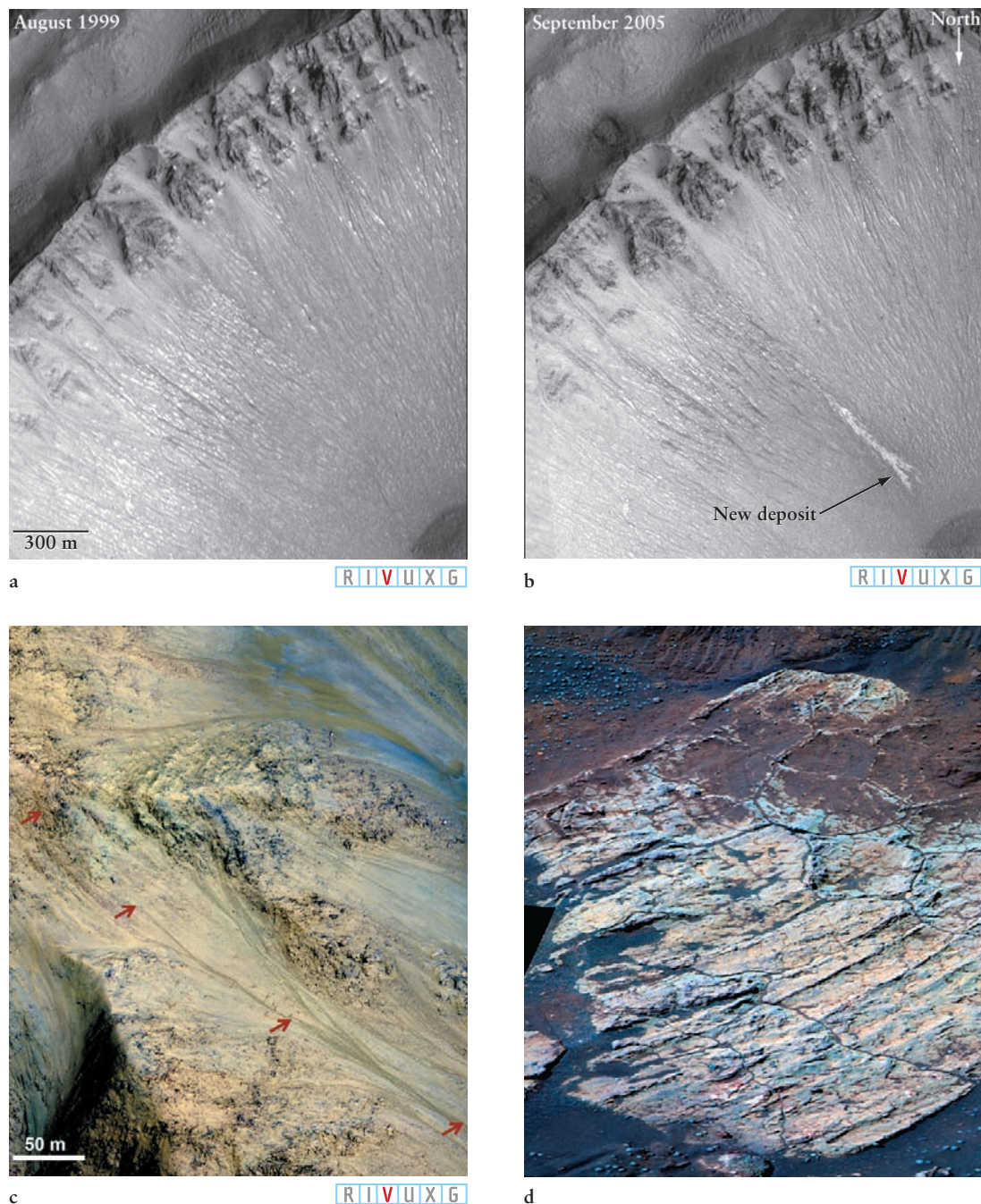


FIGURE 6-71 Martian Gullies (a, b) Two images of the same southern hemisphere crater on Mars taken 6 years apart. Whether the new deposit was initiated by ground vibration or by a short period of liquid water flow from inside Mars is still under investigation. (c) The arrows in this picture, taken of a crater in the Valles Marineris, show streaks that fade away in the winter. This may

be an indication of water flow. (d) Polygonal cracks are visible in this *Opportunity* image of Escher Rock. They were believed to have formed when this area was flooded. Water seeped into the rock, which cracked as the water froze, expanded, and then sublimated away. (a, b: NASA/JPL/Malin Space Science Systems; c: NASA/JPL-Caltech/Univ. of Arizona; d: NASA/JPL/Cornell)

crater wall is so steep that dry sandy material could have begun sliding due to a small Marsquake (earthquake on Mars) or a tremor created by a meteorite impact. *Mars Reconnaissance Orbiter* has discovered that Mars is being struck by at least 200 pieces of space debris that create craters at least 3.9 m (12.8 ft) across each year. In 2014, astronomers reported observations of streaks that change color seasonally (Figure 6-71c) that may be the first signs of active water on the red planet. The water can be liquid even in the cold temperatures of Mars's summer if it is salty, which lowers its melting point.

In 2008, astronomers using the *Mars Reconnaissance Orbiter* observed avalanches in progress close to Mars's north pole, where liquid water is even less likely to exist near the surface today (Figure 6-72). Another example of change is Mars's sand dunes, which have been observed to vary as a result of winds acting on them.

In 2004, the rover *Opportunity* discovered Martian rock (Figure 6-71d) with polygon-shaped cracks similar to dried mud flats found on Earth. The best explanation

for these features on Mars is that long after the lakes and oceans dried there, an impact released water from inside the planet that briefly reflooded the surface rock. The water seeped into the rock, and upon freezing, the water expanded and cracked the rock. This scenario would require vast amounts of water inside Mars for several reasons. First, most of the liberated water heated by the impact would have left as vapor. Second, liquid water on the surface would not persist very long because of the low air pressure. When air pressure is very low, molecules easily escape from a liquid's surface, which would cause water to vaporize. Because the pressure of Mars's atmosphere is only 0.6% that of Earth's atmosphere, any liquid water on Mars today would quickly transform into ice or furiously boil and evaporate into the thin Martian air and then be lost into space. Thus, if liquid water did cause the rock formations seen by *Opportunity*, there must have been a lot of it to start with, in order for there to be enough to cover the large areas of the Martian surface that have such cracks.

The total amount of water that once existed on Mars's surface is not known. However, by studying flood channels that exist there, geologists estimate that there was once enough water to cover the planet to a depth of 500 m (1500 ft). As noted earlier, Earth has enough water to cover our planet to a depth of 2.8 km (1.7 mi), assuming that Earth's surface was everywhere a uniform height.

Further evidence that water once flowed on Mars comes from so-called *SNC meteorites* found on Earth (Figure 6-73a). These space rocks are believed to have once been pieces of Mars because their chemistries are consistent with those of rocks studied on Mars's surface and because they contain trace gases in relative amounts found only in the current Martian atmosphere. The meteorites were ejected into space during especially powerful impacts on that planet's surface. They also contain water-soaked clay, which is not expected to be found on any objects in the solar system besides Mars, Earth, and Jupiter's moons Europa, Ganymede, and Callisto. At least 132 meteorites from Mars have been identified.

Some of the water that was on Mars is still near the surface today, frozen in several places on the red planet from its poles to its equator. Recent satellite measurements indicate that the ice at the north pole is mostly water ice and extends down about 2.0 km (1.2 mi). The south pole has layers of water ice and dry ice (frozen carbon dioxide) in roughly equal amounts. The water ice is atop the dry ice there. The temperature at Mars's poles, typically 160 K (−70°F), keeps the water permanently frozen, like the permafrost found in Siberia and Antarctica. Seasonal variations at Mars's poles (see Figure 6-63) are created by the freezing and sublimating of dry ice, which reaches a thickness of 2.0 m (2.2 yd). Indeed, during the winter, about one-third of the carbon dioxide in that hemisphere of Mars's atmosphere precipitates down as dry-ice snow.

The 1997 visit by *Pathfinder* and its little rover (see Figure 6-64b) to the Ares Vallis revealed evidence of

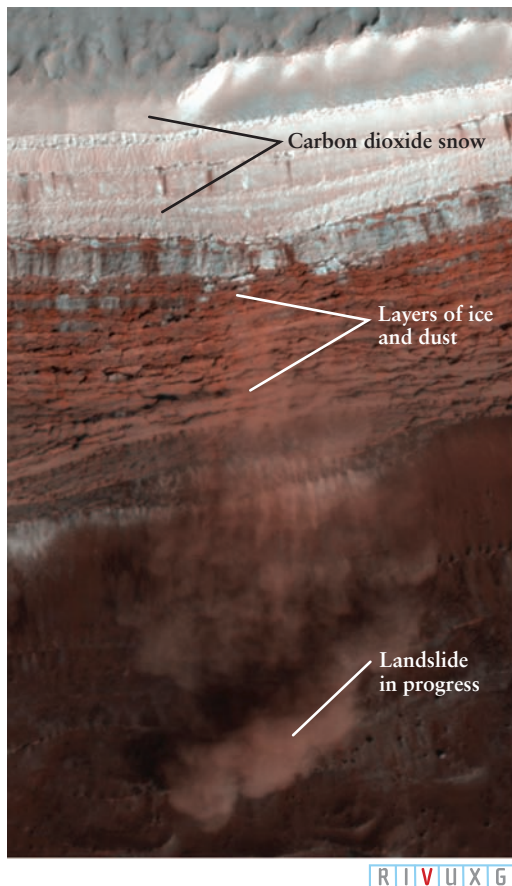


FIGURE 6-72 An Active Avalanche on Mars This image, taken by the *Mars Reconnaissance Orbiter*, shows an active landslide of falling ice, dust, and rock. This debris first hit the cliff wall about halfway down and then continued falling down the less steep slope, as seen in the bottom of the image. The cloud of debris near the bottom is about 180 m (590 ft) wide. (NASA/JPL-Caltech/University of Arizona)



FIGURE 6-73 A Piece of Mars on Earth Mars's thin atmosphere does little to protect it from impacts. Some of the debris ejected from impact craters there apparently traveled to Earth. **(a)** An SNC meteorite recovered in Antarctica. The 4.1-billion-year-old meteorite, designated ALH84001, shows



strong evidence of having been exposed to liquid water on Mars, perhaps for hundreds of years. **(b)** Possible fossil remains of primitive bacterial life on Mars, although theories of nonbiological origins have also been presented. (a: NASA; b: NASA/JPL)

several floods at the mouth of a dried flood plain. This evidence includes layers of sediment, clumps of rock and sand stuck together like similar groupings created by water on Earth, rounded rocks worn down as they were dragged by the water, and rocks aligned by water flow. The rover performed some 20 chemical analyses of rocks in the landing area. The soil at the site is a rusty color, laden with sulfur. The source of this sulfur is a mystery, as is the large amount of silicon-rich rock the rover explored.

There is observational evidence of an ancient shoreline encircling the Vastitas Borealis of Mars. The belief that the northern regions were an ocean long ago is supported by the analysis of a 1.2-billion-year-old meteorite from Mars. This rock has remnants of salt believed to have been deposited there when the Martian ocean dried up.

Dried remains of underground water channels have also been discovered on Mars. Radar signals from the *Mars Reconnaissance Orbiter* regularly penetrate Mars's surface to a depth of 1 km (0.6 mi). The return signals are processed into images showing that under the surface are concentrations of rock that appear to have been islands shaped by water flow around them. The depths of these features suggest that they were formed within the past 500 million years. The underground source of the water and the reason it flowed are still under investigation.

6-23 Search for microscopic life on Mars continues

The discovery in the 1960s that water once existed on the surface of Mars rekindled speculation about Martian life. Although it was clear that Mars has neither civilizations nor fields of plants, microbial life-forms still seemed possible. Searching for signs of organic matter was one of the main objectives of the ambitious and highly successful *Viking* missions.

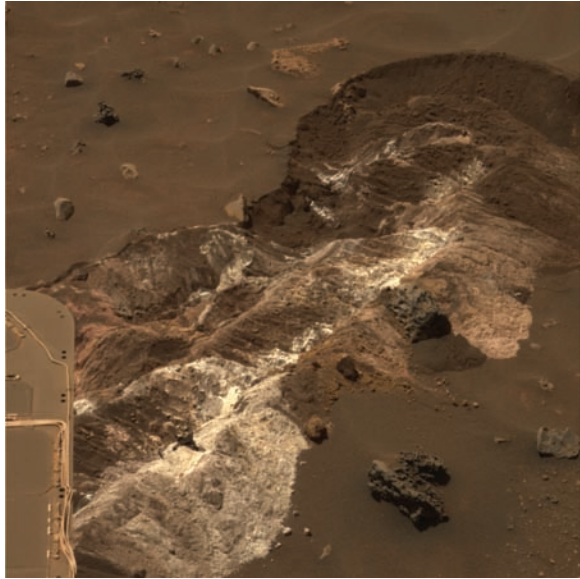
The two *Viking* spacecraft were launched during the summer of 1975. Each consisted of two modules—an orbiter and a lander. Almost one year later, both *Viking* landers set down on rocky plains north of the Martian equator.

The landers confirmed the long-held suspicion that the red color of the planet is due to large quantities of iron in its soil. Despite the high iron content of its crust, Mars has a lower average density of 3950 kg/m^3 (3.0 tons/yd^3) than that of other terrestrial planets (this value is more than 5000 kg/m^3 (3.8 tons/yd^3) for Mercury, Venus, and Earth). Mars must therefore contain overall a smaller percentage of iron than these other planets.

Each *Viking* lander was able to dig into the Martian regolith and retrieve rock samples for analysis, which showed that the rocks at both sites are rich in iron, silicon, and sulfur. The Martian regolith (**Figure 6-74**) ranges in consistency from clay to sand, although its chemistry is different from clay and sand on Earth. In 2012, *Curiosity* analyzed regolith on another part of Mars, finding it to contain olivine, feldspar, and pyroxene, which are often found in volcanic soil on Earth that has undergone extensive weathering.

The *Viking* landers each carried a compact biological laboratory designed to test for microorganisms in the Martian soil. Three biological experiments were conducted, each based on the idea that living things alter their environment: They eat, they breathe, and they release waste products.

During each experiment, a sample of the Martian regolith was placed in a closed container, with or without a nutrient substance. The container was then examined for any changes in its contents. For both landers, two of the experiments gave negative results for life. The third, however, caused great excitement. In this experiment, small amounts of water and nutrients were added



Sulfur salts

R I V U X G

FIGURE 6-74 Regolith of Mars The rover *Spirit*'s wheels churned up the regolith of Mars, revealing sulfur-based salts just below the surface. (NASA/JPL/Cornell)

and rapid and extensive changes were detected inside the sealed containers. However, further analysis showed that these changes were due solely to nonbiological chemical processes. Apparently, the Martian regolith is rich in chemicals that effervesce (fizz) when moistened. A large amount of oxygen is tied up in the regolith in the form of unstable chemicals called *peroxides* and *superoxides*, which break down in the presence of water to release oxygen gas. Complicating matters was the discovery in 2010 that the heat in the sealed containers on the *Viking* landers may have been so high that the Martian regolith deposited in them was chemically changed before it was analyzed, so that the results neither prove nor disprove the presence of organic compounds in that material.

The chemical reactivity of the Martian regolith comes from the Sun's ultraviolet radiation and from electric activity inside dust devils. Ultraviolet photons easily break apart molecules of carbon dioxide (CO_2) and water vapor (H_2O) by knocking off oxygen atoms, which then become loosely attached to chemicals in the regolith. Ultraviolet photons also produce highly reactive ozone (O_3) and hydrogen peroxide (H_2O_2) near the planet's surface, which become incorporated in the regolith. Dust devils suck up lots of dust from the planet's surface and transform it into a variety of compounds.

8 Here on Earth, hydrogen peroxide is commonly used as an antiseptic. When you pour this liquid on a wound, it fizzes and froths as the loosely attached oxygen atoms chemically combine with organic material and destroy germs. The *Viking* landers may have failed to detect any organic compounds on Mars because the superoxides and peroxides in the Martian regolith act as antiseptics there, making it sterile. Keep in mind, however, that

back when Mars had substantial amounts of water, the surface chemistry was probably substantially different and potentially more supportive of life, if it ever existed there. This is supported by the discovery in 2010 of rocks that contain carbonate that was deposited in wet, nonacidic conditions. This substance dissolves in acid-rich water. In 2013, the rover *Curiosity* drilled into the rock on Mars and analyzed the resulting powdered rock. While no life was found, carbon, nitrogen, oxygen, sulfur, phosphorus, and hydrogen, important elements for the existence of life, were detected. Furthermore, these data were taken from what had once been a wet environment that would have been chemically hospitable to life.

In 2003, astronomers began detecting methane in Mars's atmosphere. This discovery is intriguing because sunlight breaks down methane within a few centuries. Therefore, the gas needs to be continuously replenished. Indeed, plumes of methane emerging from the planet's interior have now been observed. One of these plumes was observed to contain 19,000 tons of the gas. This simple compound, CH_4 , is often a by-product of biological activity (cows come to mind in this regard). Methane is also trapped inside planets, like Earth and possibly Mars, during their formation and thereafter leaks out. In 2013, the *Curiosity* rover repeated the search for methane but found none at all. Reconciling these contradictory observations is a work in progress.

Possible signs of ancient life on Mars have been discovered here on Earth. When cut open, several of the Martian meteorites have shown microscopic features that could be fossils of Martian bacterial life and their excretions (see Figure 6-73b). These features, no larger than 500 nm (2×10^{-4} inches or 1/100 the diameter of a human hair), are 30 times smaller than bacteria found on Earth. Detailed analysis of the meteorites in 2001 revealed the presence of several organically created features, including tiny spheres similar to those found with some Earth bacteria, and magnetite crystals, which are also used by some bacteria on Earth as compasses to find food. The debate as to whether the features in these meteorites are fossil evidence of life on Mars has raged unabated for several years, and it is not yet resolved. (It is worth noting that similar claims of fossils in meteorites from Mars in the 1960s have been disproved.)

The search for microscopic life on Mars continues. If there is liquid water under the Martian surface, then it is entirely possible that life has evolved there and may still exist. This belief is based on the incredibly wide range of places that life has developed in water on Earth, from volcanic vents on the bottoms of oceans to geysers and to lakes under the Antarctic polar ice cap.

6-24 Mars's two moons look more like potatoes than spheres

Two tiny, irregularly shaped moons orbit close to Mars's surface. *Phobos* (Greek, meaning "fear") and *Deimos* ("panic") are so small that they were not discovered



a

RITVUXIG

FIGURE 6-75 Phobos and Deimos (a) Phobos, the larger of Mars's two moons, is potato-shaped and measures approximately $28 \times 23 \times 20$ km ($17 \times 14 \times 12$ mi). It is dominated by crater Stickney (on the right of the image), named for discoverer



b

RITVUXIG

Asaph Hall's wife (Angeline Stickney). (b) Deimos is less cratered than Phobos and measures roughly $16 \times 12 \times 10$ km ($10 \times 7 \times 6$ mi). (a and b: NASA/JPL-Caltech/University of Arizona)

until 1877. In mythology, these names are either Mars's sons by Aphrodite or the two horses that pulled Mars's war chariot—take your pick. Potato-shaped Phobos is the inner and larger of the two moons (Figure 6-75a). It is heavily cratered, and its surface has been transformed into dust at least 1.0 m (1.1 yd) thick by countless tiny impacts over the eons. Deimos is smaller and less cratered than Phobos (Figure 6-75b).

The gravitational deflections by Phobos of spacecraft passing near it have been used to determine the moon's mass. Measurements of that mass and its volume, made in 2010, reveal a very low density (density = mass/volume), which suggests that Phobos may be extremely porous, with up to 30% of its interior composed of empty space.

Both Phobos and Deimos orbit in the same direction that Mars rotates, just as our Moon orbits Earth. However, Phobos is so close to the planet that this moon goes around 3 times in 1 Martian day. This rapid revolution creates the unusual situation wherein Phobos rises in the west, races across the sky in only $5\frac{1}{2}$ hours as seen from Mars's equator, and sets in the east. As seen from Mars, Deimos rises in the east and takes about 3 Earth days to creep from one horizon to the other.

Phobos creates tiny tides on Mars, which in turn pull that rapidly orbiting moon inward (just the opposite effect of Earth on our more slowly orbiting Moon). Phobos, descending at a rate of 1.8 m/Earth year (2.0 yd/Earth year), will either strike Mars or be pulled apart by the tidal force and become a ring within 50 million years.

Insight Into Science

Imagine the Moon The definitions we use for words based on our everyday experience often fail us in astronomy. For example, the word *moon* usually creates an image of a spherical body, like our Moon. In reality, most of the moons in the solar system have very irregular shapes, like Phobos and Deimos.

Phobos and Deimos were not formed like our Moon, which condensed from debris that “splashed” into orbit as the result of a collision between our planet and another large body. For decades Phobos and Deimos were both believed to be captured asteroids (which may still be true for Deimos). While some asteroids have densities as low as that of Phobos, astronomers have trouble showing how Phobos could have been captured and put into its relatively circular orbit without being pulled apart. A viable explanation of Phobos's presence is still under development. Both moons are in synchronous rotation as they orbit Mars.

COMPARATIVE PLANETOLOGY OF THE INNER PLANETS

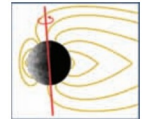
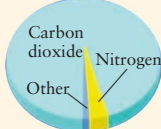
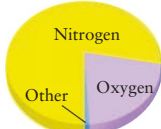
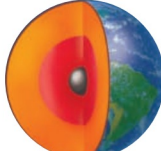
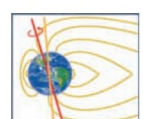
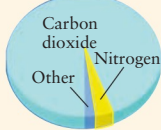
6-25 Comparisons of planetary features provide new insights

Now that we have examined the terrestrial planets individually, it can be useful to see how their various features compare to each other and to Earth. **Table 6-4**,

Focus Question 6-20

Would our Moon have to be closer or farther away to orbit in the same direction but rise in the west?

TABLE 6-4 THE INNER PLANETS: A COMPARISON

Size and Mass	Atmosphere	Surface	Interior	Rotation and Magnetic Field
Mercury  38% of Earth's diameter A scant 5.5% of Earth's mass Much closer in size to Earth's Moon than to the other terrestrial planets	Thinnest of all the terrestrial planets; H ₂ , He, K, Na, O ₂ , Ca, and Mg  700 K (800°F) by day, 100 K (−280°F) by night Mercury's gravity is too low to hold any gases as a permanent atmosphere.	Heavy cratering, scarps Extinct volcanoes No known internal activity, so surface remains unchanged except for cratering No surface water	Iron-nickel core, rocky shell (details unknown) 	1 Mercury solar day = 176 solar Earth days A day on Mercury is 2 Mercurian years long! Weak global field 
Venus  Nearly 95% of Earth's diameter 82% of Earth's mass Almost the sister planet to Earth in mass, size, and composition	Primarily carbon dioxide, with a minor component of nitrogen; 750 K (800°F)  Carbon dioxide 90x denser than Earth's atmosphere Venus's thick atmosphere creates a greenhouse effect; protects surface from all but the most massive infalling space debris.	Light cratering, hills, both extinct and active volcanoes Most craters melted or covered by magma Like Earth, continentlike plateaus and ocean bottomlike lowlands No surface water	Iron-nickel core, rocky shell (details unknown) 	1 Venus solar day = 117 Earth solar days No magnetic field 
Earth  Largest and most massive of all four terrestrial planets	Primarily nitrogen and oxygen; 200–315 K  Earth's surface temperature ranges from about 330 K (140°F) to 183 K (−130°F)	Very little cratering, continents and land at ocean floors, weathering, both extinct and active volcanoes, global tectonic plates Earth has removed most of its craters by tectonic plate motion and weathering. As with Venus, Earth's atmosphere protects the surface from most impacts.	Iron-nickel core, rocky shell 	Earth's solar day is shortest at 24 h Moderate global field A liquid iron core, combined with Earth's rotation, creates a strong magnetic field. 
Mars  About half (53%) of Earth's diameter 11% of Earth's mass	Carbon dioxide 0.006 times as dense as Earth's atmosphere; 280 K (45°F) to 160 K (−170°F)  The thin atmosphere has enabled many pieces of space debris to strike the planet and form craters	Moderate cratering, weathering, dormant volcanoes, huge canyons No surface water; large volumes of frozen water and frozen carbon dioxide just below surface, mostly near poles Continual erosion of surface, primarily by wind Surface from its North Pole half way to its equator is lower than the rest of the planet's surface No real continentlike regions	Iron-nickel core, rocky shell (details unknown) 	Virtually the same solar day as Earth (24 h 39 min) Its molten iron core is much smaller than that of Earth and it has no global magnetic field, only local magnetic fields. 

The Inner Planets: A Comparison, summarizes much of this material.

SUMMARY OF KEY IDEAS

Earth: A Dynamic, Vital World

- Earth's atmosphere is about four-fifths nitrogen and one-fifth oxygen. This abundance of oxygen is due to the biological processes of plant life absorbing carbon dioxide and giving off oxygen as a waste product. Earth's atmosphere is divided into layers named the troposphere, stratosphere, mesosphere, and ionosphere. Ozone molecules in the stratosphere absorb ultraviolet light rays.
- Earth's surface is divided into huge plates that move over the upper mantle. Movement of these plates, a process called *plate tectonics*, is caused by convection in the mantle. Also, upwelling of molten material along cracks in the ocean floor occurs during seafloor spreading. Plate tectonics is responsible for most of the major features of Earth's surface, including mountain ranges, volcanoes, and the shapes of the continents and oceans.
- Study of seismic waves (vibrations produced by earthquakes) shows that Earth has a small, solid inner core surrounded by a liquid outer core. The outer core is surrounded by the dense mantle, which in turn is surrounded by the thin, low-density crust on which we live. Earth's inner and outer cores are composed primarily of iron. The mantle is composed of iron-rich minerals.
- Earth's magnetic field produces a magnetosphere that surrounds the planet and deflects the solar wind. Some charged particles from the solar wind are trapped in two huge, doughnut-shaped rings called the Van Allen radiation belts. An Earthward deluge of particles from a coronal mass ejection on the Sun can pierce the belts and produce exceptional auroras.

The Moon and Tides

- The Moon has heavily cratered highlands and relatively smooth-surfaced maria.
- Impacts have been the only significant “weathering” agent on the Moon; the Moon's regolith (pulverized rock layer) was formed by meteoritic action.
- Powdered into regolith, the anorthosite rock of the highland is brighter than the powdered basalts of the maria.
- Many lunar rock samples are solidified lava formed largely of minerals also found in Earth rocks.
- Anorthositic rock in the lunar highlands was formed between 4.0 and 4.3 billion years ago, whereas the mare basalts solidified between 3.1 and 3.8 billion years ago. The Moon's surface has undergone very little geologic change over the past 3 billion years.
- Frozen water has been discovered in numerous places just below the Moon's surface.
- The collision-ejection theory of the Moon's origin, accepted by most astronomers, holds that the young Earth

was struck by a huge planetesimal, and debris from this collision coalesced to form the Moon.

- The Moon was molten in its early stages, and the anorthositic crust solidified from low-density magma that floated to the lunar surface. The mare basins were created later by the impact of planetesimals and were then filled with lava from the lunar interior.
- Gravitational interactions between Earth and the Moon produce tides in the oceans of Earth and set the Moon into synchronous rotation. The Moon is moving away from Earth and, consequently, Earth's rotation rate is decreasing.

Mercury

- All four inner planets are composed primarily of rock and metal, and thus they are classified as terrestrial.
- Even at its greatest orbital elongations, Mercury can be seen from Earth only briefly after sunset or before sunrise.
- The Mercurian surface is pocked with craters like the Moon's, but extensive, smooth plains lie between these craters. Long cliffs meander across the surface of Mercury. These scarps formed as the planet cooled, solidified, and shrank.
- The long-ago impact of a large object formed the huge Caloris Basin on Mercury and shoved up jumbled hills on the opposite side of the planet. Several similar but smaller impact features also exist on Mercury and on our Moon.
- Mercury has an iron core, which fills more of its interior than Earth's core fills Earth.
- Mercury has a weak, global magnetic field that partially shields it from the solar wind.

Venus

- Venus is similar to Earth in size, mass, and average density, but it is covered by unbroken, highly reflective clouds composed primarily of sulfuric acid that conceal its other features from observers using visible-light telescopes.
- Although most of Venus's atmosphere is carbon dioxide, its clouds are composed of sulfuric acid mixed with yellowish sulfur dust. Active volcanoes on Venus are likely to be a constant source of this sulfurous veil.
- Venus's exceptionally high temperature is caused by the greenhouse effect, as the dense carbon dioxide atmosphere traps and retains heat emitted by the planet. The surface pressure on Venus is 92 atm, and the surface temperature is 750 K (890° F). Both temperature and pressure decrease as altitude increases.
- The surface of Venus is surprisingly flat and mostly covered with gently rolling hills. It has very few impact craters compared to Mercury or our Moon. There are two major “continents” and several large volcanoes. The surface of Venus shows evidence of local tectonic

activity but not the large-scale motions that play a major role in continually reshaping Earth's surface.

Mars

- Earth-based observers found that the Martian solar day is nearly the same as that of Earth, that Mars has polar ice caps of carbon dioxide snow that expand and shrink with the seasons, and that the Martian surface undergoes seasonal color changes.
- The Martian surface has many flat-bottomed craters, several huge volcanoes, a vast equatorial canyon, and dried-up riverbeds—but no canals formed by intelligent life. River deltas and dry riverbeds on the Martian surface indicate large amounts of water once flowed there.
- Liquid water would quickly boil away in Mars's thin present-day atmosphere, but the planet's polar ice caps contain significant quantities of frozen water, and a layer of permafrost exists beneath parts of the regolith.
- The Martian atmosphere is composed mostly of carbon dioxide. The surface pressure is less than 0.01 atm.
- Mars has no global magnetic fields, but local fields pierce its surface in many places.
- Mars has two irregularly shaped moons, Phobos and Deimos. Both are in synchronous rotation with Mars. How they came into orbit is still under investigation.

WHAT DID YOU THINK?

- 1 *Can Earth's ozone layer, which has been partially depleted, be naturally replenished?* Yes. Ozone is created continuously from normal oxygen molecules by their interaction with the Sun's ultraviolet radiation.
- 2 *Do we see all parts of the Moon's surface at some time during the lunar cycle?* No. Because the Moon's rotation around Earth is synchronous, we always see the same side. The far side of the Moon has been seen only from spacecraft that pass or orbit it.
- 3 *Does the Moon rotate and, if so, how fast?* The Moon rotates at the same rate that it revolves around Earth, once every 27.3 Earth days. If the Moon did not rotate, then, as it revolved, we would see its entire surface from Earth, which we do not.
- 4 *What causes the ocean tides?* The tides are created by gravitational forces from the Moon and Sun combined with Earth's motion around the barycenters between it and those other two bodies.
- 5 *Which terrestrial planet—Mercury, Venus, Earth, or Mars—has the coolest surface temperature?* The nighttime side of Mercury, closest planet to the Sun, is the coldest surface of any terrestrial planet.
- 6 *Which planet is most similar in size to Earth?* Venus is most similar to Earth in size.

7 *Which terrestrial planet—Mercury, Venus, Earth, or Mars—has the hottest surface temperature?* Venus is hottest, its temperature raised above that of Mercury because of the greenhouse effect in Venus's atmosphere.

8 *Is life known to exist on Mars today?* No current life has yet been discovered on Mars, but it may exist in underground water oceans.

Review Questions

1. What type of chemical or molecule is most common in Earth's atmosphere?
 - a. carbon dioxide
 - b. oxygen
 - c. water
 - d. nitrogen
 - e. hydrogen
2. Why does Earth's albedo change daily? Seasonally?
3. Why is Earth's surface not riddled with craters as is that of the Moon?
4. List the layers of Earth's atmosphere.
5. Describe the process of plate tectonics. Give specific examples of geographic features created by plate tectonics.
6. How do we know about Earth's interior, given that the deepest wells and mines extend only a few kilometers into its crust?
7. What occurs on Earth when two tectonic plates of roughly equal density collide head-on?
8. What occurs when a higher-density tectonic plate collides head-on with a lower-density plate?
9. Describe the interior structure of Earth.
10. Why is the center of Earth not molten?
11. What are the Van Allen radiation belts?
12. Why do charged particles from the Sun (the solar wind) get trapped in the Van Allen radiation belts?
13. The Moon's surface is best described as:
 - a. fine-grained powder
 - b. solid rock
 - c. rocky rubble
 - d. liquid water oceans and dry land
 - e. molten rock
14. Explain why the maria appear darker than the lunar highlands.
15. Why are there so few craters on the maria?
16. Where on the Moon did the largest impact craters occur?
17. How do we know that the maria were formed after the lunar highlands?
18. Describe how the Moon creates ocean tides on Earth.

19. What is the difference between spring tides and neap tides? During which phase(s) of the Moon does each occur?
20. Why do most scientists support the collision-ejection theory for the Moon's formation?
21. Why are nearly all of the craters on the Moon circular, even though many impacts there were not head-on?
22. Why hasn't the water ice near the Moon's poles been evaporated by sunlight?
23. Which pair of planets has atmospheres with the most similar chemical compositions?
 - a. Earth and Venus
 - b. Earth and Mars
 - c. Venus and Mars
 - d. Mercury and Mars
 - e. Mercury and Venus
24. Which planet is least likely to have water ice on or just under its surface?
 - a. Earth
 - b. Mercury
 - c. Mars
 - d. Venus
25. Which object is most similar to Venus in mass and diameter?
 - a. Earth
 - b. Mars
 - c. Mercury
 - d. our Moon
26. Compare the surfaces of Mercury and our Moon. How are they similar? How are they different?
27. Compare the interiors of Mercury and Earth. How are they similar? How are they different?
28. On Mercury, when would you expect it to be coldest: noon, just before sunset, midnight, or just before sunrise? Explain your reasoning.
29. Why is Mercury more heavily cratered than Earth?
30. Astronomers often refer to Venus as Earth's twin. What physical properties do the two planets have in common? In what ways are the two planets dissimilar?
31. Why is it hotter on Venus than on Mercury?
32. What evidence exists for active volcanoes on Venus?
33. Describe the Venusian surface. What kinds of geologic features would you see if you could travel around the planet?
34. Why is Mars red?
35. What is the greenhouse effect? What role does it play in the atmospheres of Venus and Earth?
36. What geologic features indicate that plate tectonic activity once occurred on Mars? What features created by tectonic activity on Earth are not found on Mars?
37. Why are volcanoes on Mars taller than those on Earth?
38. What evidence have astronomers accumulated that liquid water once existed in large quantities on Mars's surface? What evidence is there that water still exists under its surface?

Advanced Questions

39. Explain how the outward flow of energy from Earth's interior drives the process of plate tectonics.
40. Why are active volcanoes, such as Mount St. Helens, usually located in mountain ranges along the boundaries of tectonic plates?
41. Why is more lunar detail visible through a telescope when the Moon is near quarter phase than when it is at full phase?
42. Why are the Moon rocks retrieved by astronauts so much older than typical Earth rocks, even though both worlds formed at nearly the same time?
43. Apollo astronauts left seismometers on the Moon that radioed seismic data back to Earth. The data showed that moonquakes occur more frequently when the Moon is at perigee (closest to Earth) than at other locations along its orbit. Give an explanation for this finding.
44. Venus takes 440 days to move from greatest western elongation to greatest eastern elongation, but it needs only 144 days to go from greatest eastern elongation to greatest western elongation. With the aid of a diagram, explain why.
45. How might Venus's cloud cover change if all of Venus's volcanic activity suddenly stopped? How might these changes affect the overall Venusian environment?
46. Compare Venus's continents with those on Earth. What do they have in common? How are they different?
47. Explain why Mars has the longest synodic period of all of the planets, although its sidereal period is only 687 days.
48. With carbon dioxide accounting for about 95% of the atmospheres of both Mars and Venus, why is there little greenhouse effect on Mars today?

Discussion Questions

49. Comment on the idea that without the Moon's presence, life would have developed far more slowly.
50. Identify and compare the advantages and disadvantages of lunar exploration by astronauts as opposed to mobile, unmanned instrument packages and robots.
51. When was the last earthquake near your hometown? How far is your hometown from a plate boundary? What kinds of topography (for example, mountains, plains, seashore) dominate the geography of your hometown area? Does that topography and the frequency of earthquakes seem to be consistent with your hometown's proximity to a plate boundary?
52. If you were planning a new *Messenger* mission to Mercury, what features and observations would be of particular interest to you and why?

53. If you were designing a space vehicle to land on Venus, what special features would be necessary? In what ways would this mission and landing craft differ from a spacecraft designed for a similar mission to Mercury?

54. Compare the scientific opportunities for long-term exploration offered by the Moon and Mars. What difficulties would there be in establishing a permanent base or colony on each of these two worlds?

55. Imagine you are an astronaut living at a base on Mars. Describe your day's activities, what you see, the weather, the spacesuit you are wearing, and so on.

Got It?

56. Will the Earth's ozone layer, which has been depleted over decades, be naturally replenished? If so, how?

57. Since infrared radiation (heat) from the Sun does not enter a greenhouse, how do greenhouses get warm?

58. How can we tell by looking at the Moon that it rotates?

59. How were the craters on the Moon formed?

60. What gas in the atmosphere of Venus is primarily responsible for the greenhouse effect there?

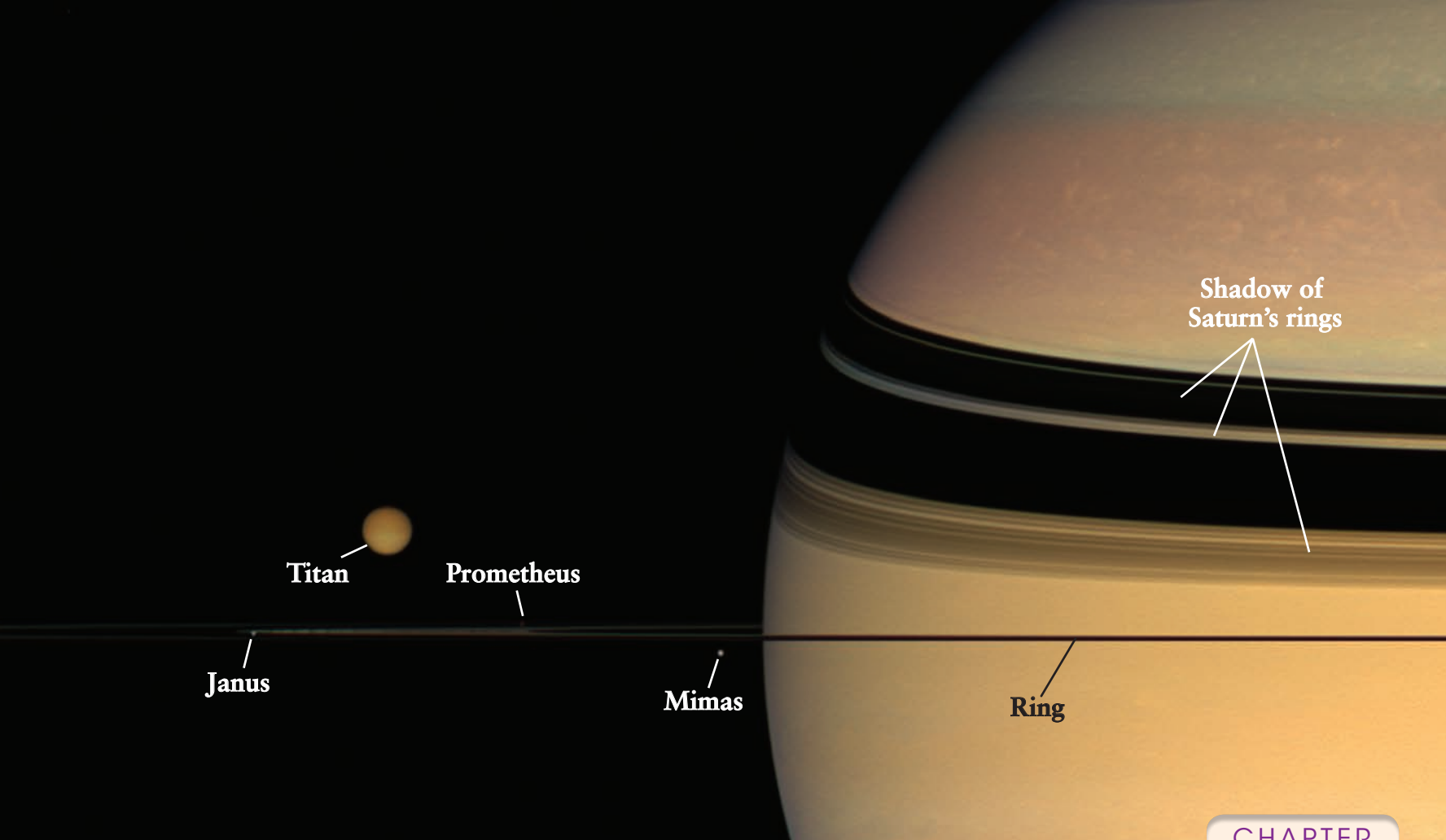
61. Which of the following statements about life on Mars is correct?

- a. Rovers from Earth have observed life on the surface of Mars.
- b. Rovers from Earth have found remnants of simple life on Mars.
- c. Rovers from Earth have found remnants of advanced civilizations on Mars.
- d. Rovers from Earth have found life underground.
- e. Rovers from Earth have not found evidence of life on Mars.

62. What color does the surface of Venus have, as seen from that world, and why?

63. Which terrestrial planet has the coldest surface, and why?

This page intentionally left blank



R I V U X I G **Saturn, Four of Its Moons, and Its Rings** This *Cassini* image gives you a feel for how thin the rings of Saturn are. (NASA/JPL/Space Science Institute)

The Outer Planets and Their Moons

WHAT DO YOU THINK?

- 1 Is Jupiter a “failed star”? Why or why not?
- 2 What is Jupiter’s Great Red Spot?
- 3 Does Jupiter have continents and oceans?
- 4 Is Saturn the only planet with rings?
- 5 Are the rings of Saturn solid ribbons?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

The four giant worlds lack solid surfaces and are so much larger, rotate so much faster, and have such different chemical compositions from our planet that upon seeing them one knows, to paraphrase Dorothy, “You’re not on Earth anymore.” There is nothing on Earth remotely similar to the swirling red and brown clouds of Jupiter, the ever-changing ring system of Saturn, or the blue-green clouds of Uranus and Neptune. The outer four worlds collectively have at least 170 moons, which have an amazing range of shapes, sizes, and surfaces. Some of these moons have water interiors; some have volcanoes; some have surface ice; one, Titan, has a thick atmosphere; and one, ice-covered Hyperion, is arguably the most bizarre-looking

object in the solar system. We begin our exploration of the outer planets, and their moons and rings, with magnificent Jupiter.

In this chapter you will discover

- **Jupiter, an active, vibrant, multicolored world more massive than all the other planets combined**
- **Jupiter’s diverse system of moons**
- **Saturn, with its spectacular system of thin, flat rings and numerous moons, including bizarre Enceladus and Titan**
- **Uranus and Neptune, ice giants similar to each other and different from the gas giants, Jupiter and Saturn**

JUPITER: Vital Statistics

Planet symbol:	♃
Average distance from the Sun:	5.20 AU = 7.78×10^8 km = 4.83×10^8 mi
Maximum distance from the Sun:	5.46 AU = 8.16×10^8 km = 5.07×10^8 mi
Minimum distance from the Sun:	4.95 AU = 7.41×10^8 km = 3.08×10^8 mi
Orbital eccentricity:	0.048
Average orbital speed:	13.1 km/s = 8.14 mi/s
Orbital period:	11.86 years
Rotation period:	9 h 50 m 28 s (equatorial) 9 h 55 m 30 s (polar = internal)
Inclination of equator to orbit:	3.12°
Inclination of orbit to ecliptic:	1.30°
Diameter:	142,984 km = 88,850 mi = 11.21 Earth diameters (equatorial) 133,700 km = 83,000 mi = 10.48 Earth diameters (polar)
Mass:	1.90×10^{27} kg = 318 Earth masses
Average density:	$1330 \text{ kg/m}^3 = 2240 \text{ lb/yd}^3 = 0.241$ Earth density
Escape speed:	60.2 km/s = 37.4 mi/s
Surface gravity (Earth = 1):	2.36
Albedo:	0.52
Average temperature at cloud tops:	−108°C = −162°F = 165 K
Atmospheric composition (by number of molecules):	89.8% hydrogen (H ₂), 10.2% helium (He), and traces of methane (CH ₄), ammonia (NH ₃), water vapor (H ₂ O), and other gases



FIGURE 7-1 Jupiter and Its Vital Statistics This view was sent back from *Voyager 1* in 1979. Features as small as 600 km (370 mi) across can be seen in the turbulent cloud tops of this giant planet. The complex cloud motions that

surround the Great Red Spot are clearly visible. Also, clouds at different latitudes have different rotation rates. The inset image of Earth shows its size relative to Jupiter. (NASA/JPL; inset: NASA)

JUPITER

Jupiter's multicolored bands, dotted with ovals of white and brown (Figure 7-1), give it the appearance of a world unlike any of the terrestrial planets. Viewed through even a small telescope (see Figure 2-11b), you can see up to four of its moons. As the high-resolution images in this chapter reveal, Jupiter is a world of breathtaking beauty.

1 Jupiter is the largest planet in the solar system: More than 1300 Earths could be packed into its volume. Using the orbital periods of its moons in Kepler's laws, astronomers have determined that Jupiter is 318 times more massive than Earth. Indeed, Jupiter has more than $2\frac{1}{2}$ times as much mass as all the other planets combined. This huge mass has created the myth that Jupiter is a failed star, meaning that it has almost enough matter to shine on its own, like the Sun, which we will study in Chapter 9. In fact, Jupiter would have to be 75 times more massive than it is to generate energy like the Sun and, therefore, to be classified as a star. Nevertheless, Jupiter emits approximately twice as much energy as it receives from the Sun. Its extra energy comes from radioactive elements in its core and from an overall contraction amounting to less than 10 cm (4 in) per century. Contraction contributes to heating because when any matter is compressed, it heats up.

7-1 Jupiter's outer layer is a dynamic area of storms and turbulent gases

Jupiter is permanently covered with clouds (see Figure 7-1). Because it rotates about once every 10 hours—the fastest of any planet—Jupiter's clouds are in perpetual motion and are confined by this motion to narrow bands of latitude. Even through a small telescope, you can see Jupiter's

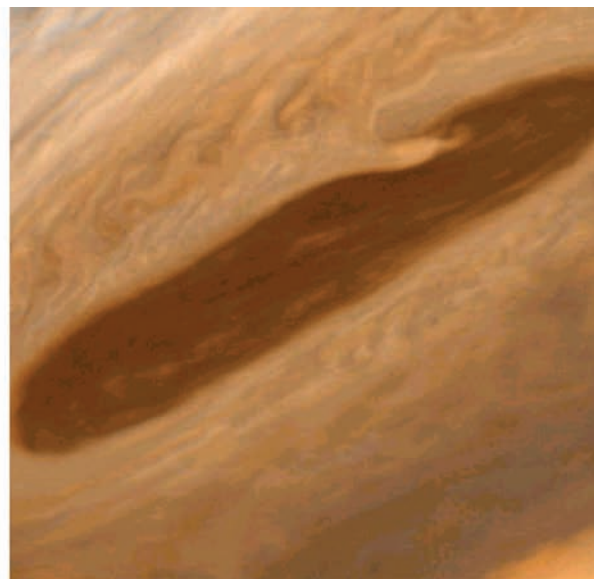
dark, reddish bands called **belts**, alternating with light-colored bands called **zones**. The zones are white due to ammonia vapor at their tops, while the sources of the colors of the belts are still under investigation. These belts and zones are gases flowing eastward or westward, with very little north–south motion. Such winds are called a *zonal flow*, and why some winds flow one way and some the other is also still under investigation. In contrast, winds on slower-rotating Earth wander over vast ranges of latitude (compare Figure 6-1). Wind speeds between 310 km/h (190 mi/h) and 645 km/h (400 mi/h) are typical on Jupiter.

Jupiter's belts and zones provide a framework for turbulent, swirling cloud patterns, as well as rotating storms similar in structure to hurricanes, cyclones, or typhoons on Earth. These storms are known as *white ovals* and *brown ovals* (Figure 7-2). The white ovals are observed to be cool clouds higher than the average clouds in Jupiter's atmosphere. The brown ovals are warmer and lower clouds, seen through holes in the normal cloud layer. The various oval features last from hours to centuries. Computers show us how the cloud features on Jupiter would look if the planet's atmosphere were unwrapped and laid flat like a map of Earth (Figure 7-3).

Jupiter's most striking feature is its **Great Red Spot** (Figure 7-4), which is so large that it can be seen through a small telescope. This feature changes size. At present it is about 16,000 km (9940 mi) from left to right by 13,000 km (8080 mi) from top to bottom—large enough for Earth (diameter 12,800 km or 7950 mi) to fit inside it. Sometimes it has been large enough for three Earths to fit in it. At other times it has been only 11,000 km (6840 mi) from left to right, smaller than the diameter of Earth. The Great Red Spot was



a



b

FIGURE 7-2 Close-Ups of Jupiter's Atmosphere The dynamic winds, rapid rotation, internal heating, and complex chemical composition of Jupiter's atmosphere create its beautiful and complex banded pattern. (a) A *Voyager 2* southern hemisphere

image showing a white oval that has existed for more than 40 years. (b) A *Voyager 2* northern hemisphere image showing a brown oval. The white feature overlapping the oval is a high cloud. (NASA)

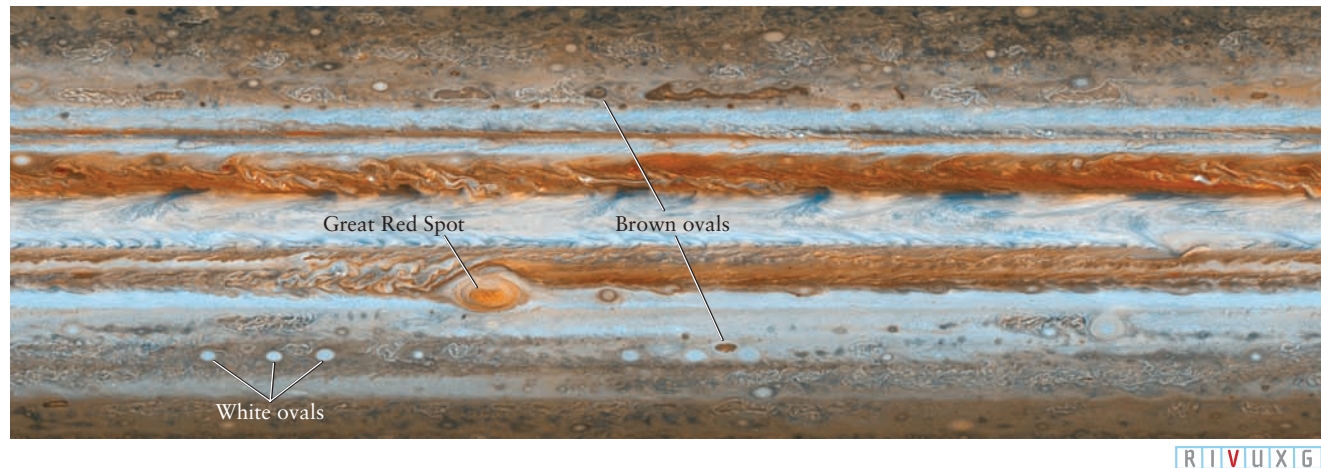


FIGURE 7-3 Jupiter Unwrapped *Cassini* images of Jupiter were combined and opened to give a maplike representation

first observed around 1656, either by the English scientist Robert Hooke or the Italian astronomer Giovanni Cassini. Because earlier telescopes were unlikely to have been able to see it, the Great Red Spot could well have formed long before that time.

2 The Great Red Spot is a storm of swirling gases much like a hurricane or typhoon. Heat welling upward from inside Jupiter has maintained it for more than three

of the planet. The banded structure is absent near the poles. (NASA/JPL-Caltech)

and a half centuries. (Consider what life would be like for us if Earth sustained storms for such long periods!) Between 1998 and 2000, three smaller white storms on Jupiter that formed in the 1930s merged (Figure 7-5a–d), creating a larger white storm that became red in 2006 (Figure 7-5e inset). Called Red Spot Jr. (technically named Oval BA), it is similar to, but somewhat smaller than, the Great Red Spot and located at nearly the same latitude (Figure 7-5e). As with the Great Red Spot, the cause of the red color is still being studied.

In 1690, Cassini noticed that the speeds of Jupiter's clouds vary with latitude, an effect called **differential rotation**. Near the poles, the rotation period of Jupiter's atmosphere, 9 h 55 min 30 s, is 5 min longer than at the equator. Furthermore, clouds at different latitudes circulate in opposite directions—some eastward, some westward. At their boundaries, the clouds rub against each other, creating beautiful swirling patterns (see Figures 7-2, 7-4, and 7-5c). The interactions of clouds at different latitudes also help provide stability for storms like the Great Red Spot.

Jupiter's rapid rotation causes its equatorial radius to be 9284 km (5770 mi) greater than its polar radius. All such rotating worlds, wider at the equator than at the poles, are said to be oblate.

3 Astronomers first determined Jupiter's overall chemical composition from its average density, which is only 1330 kg/m^3 (2240 lb/yd^3). Recall from Section 6-1 that average density is mass divided by volume and, as noted there, we use lb/yd^3 for density in US customary units. Please don't confuse this usage of lb with its normal usage as weight. We determine Jupiter's mass using Kepler's third law and the orbital periods of its moons, while the trigonometry of Jupiter's distance from Earth and angular size in our sky reveal its diameter and, hence, its volume. This low density implies that Jupiter is composed of primarily the lightweight elements hydrogen and helium surrounding a relatively small central volume thousands of kilometers below its cloud tops containing water, metal, and rock. The planet has no solid continents, islands, or water oceans on

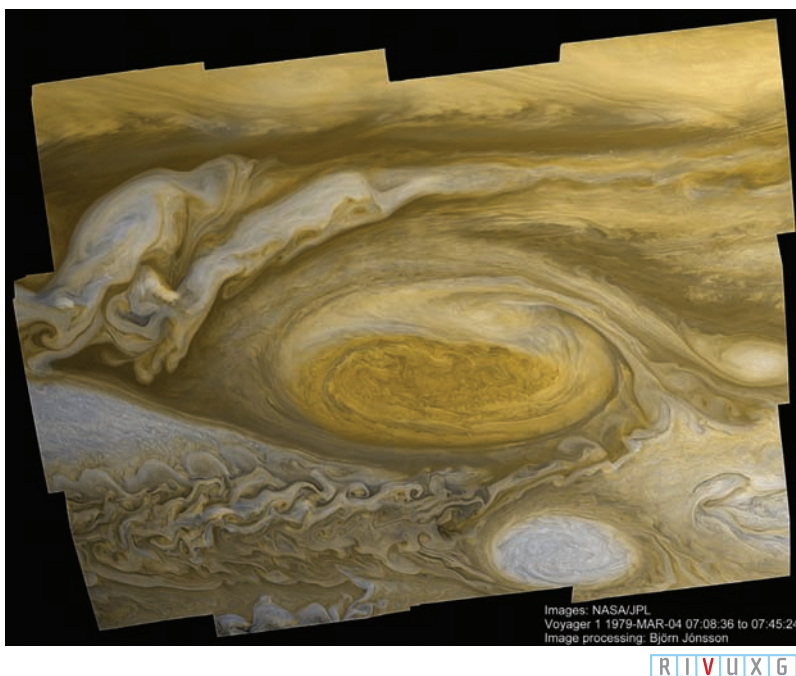


FIGURE 7-4 The Great Red Spot This image of the Great Red Spot taken in 2011 shows the counterclockwise circulation of gas in the Great Red Spot that takes about 6 days to make one rotation. The clouds that encounter the spot are forced to pass around it, and when other oval features are near it, the entire system becomes particularly turbulent, like batter in a two-bladed blender. (NASA/JPL; Digital processing: Björn Jónsson)

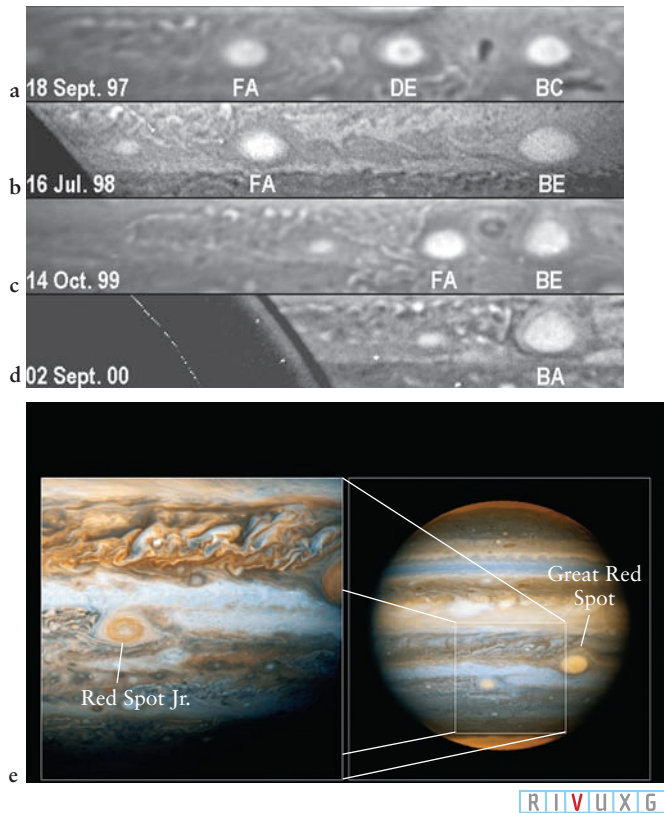


FIGURE 7-5 Creating Red Spot Jr. (a–d) For 60 years prior to 1998, the three white ovals labeled FA, DE, and BC traveled together at the same latitude on Jupiter. Between 1998 and 2000, they combined into one white oval, labeled BA, which (e) became a red spot, named Red Spot Jr., in 2006. (a–d: NASA/JPL/WFPC2; e: NASA, ESA, A. Simon-Miller [NASA/GSFC], and I. de Pater [University of California Berkeley])

its surface. Because their primary elements are hydrogen and helium, Jupiter and Saturn are called *gas giants*.

Spectra from Earth-based telescopes and from the *Galileo* probe sent into Jupiter's upper atmosphere in 1995 give more detail about the chemistry of this giant world's upper level. About 86% of its atoms are hydrogen and 13% are helium. The remainder consists of molecular compounds, such as methane (CH_4), ammonia (NH_3), and water vapor (H_2O). Keeping in mind that different elements have different masses, we can convert these percentages of atoms into the masses of various substances in Jupiter's atmosphere: 75% hydrogen, 24% helium, and 1% other substances. Because the interior contains heavier elements than the surface, the overall mass distribution in Jupiter has been calculated to be 71% hydrogen, 24% helium, and 5% heavier elements.

Observations from spacecraft visiting Jupiter and its moons, combined with the scientific model of Jupiter's atmosphere developed to explain the observations of Jupiter's clouds and chemistry, indicate that it has three major cloud layers (Figure 7-6a). The uppermost Jovian cloud layer is composed of crystals of frozen ammonia. These crystals and the frozen water in Jupiter's clouds are white, so what chemicals create the subtle tones of brown, red, and orange?

The answer is as yet unknown. Some scientists think that sulfur compounds, which can assume many different colors depending on their temperature, play an important role; others think that phosphorus is involved, especially in the Great Red Spot. The middle cloud layer is primarily ammonium hydrosulfide, and the bottom cloud layer is mostly composed of water vapor.

The descent of a probe from the *Galileo* spacecraft into Jupiter's atmosphere in 1995 revealed wind speeds of up

to 600 km/h (375 mi/h), higher-than-expected air density and temperature, and lower-than-expected concentrations of water, helium, neon, carbon, oxygen, and sulfur. These seemingly abnormal observations probably occurred because the probe descended into a particularly arid region of the atmosphere called a *hot spot*, akin to the air over a desert on Earth.

Focus Question 7-1

What are the two most common elements in Jupiter's upper layers?

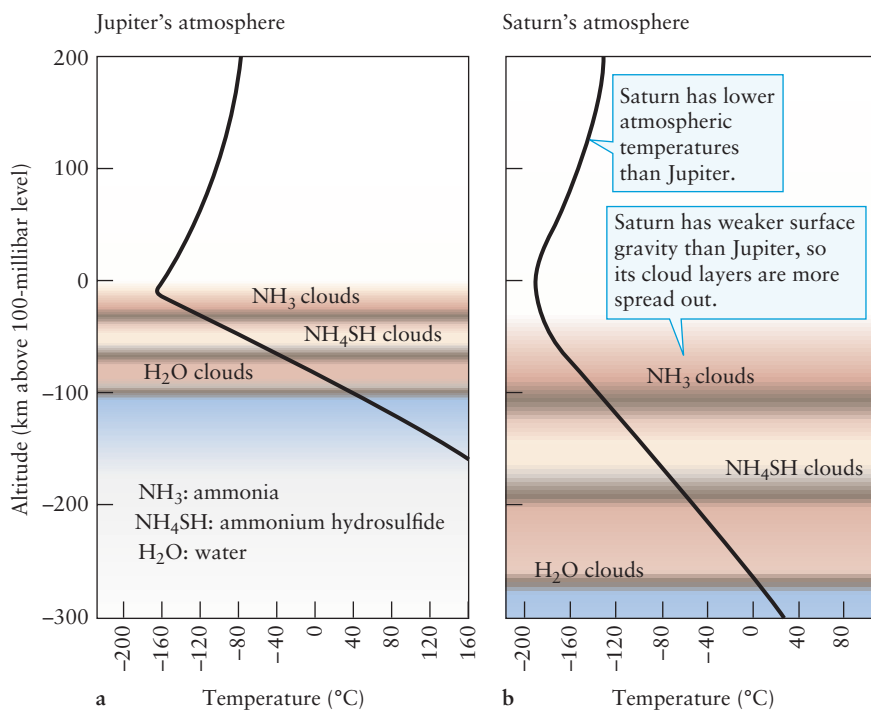


FIGURE 7-6 Jupiter's and Saturn's Upper Layers These graphs display temperature profiles of (a) Jupiter's and (b) Saturn's upper regions, as deduced from measurements at radio and infrared wavelengths. Three major cloud layers are shown in each, along with the colors that predominate at various depths. Data from the *Galileo* spacecraft indicate that Jupiter's cloud layers are not found at all locations around the planet; there are some relatively clear, cloud-free areas.

Below its cloud layer, Jupiter's mantle is entirely liquid. Here on Earth, the distinction between the gaseous air and the liquid oceans is very clear—jump off a diving board and you know when you hit the water. However, the conditions on Jupiter under which hydrogen liquefies are different from anything we normally experience. Solid, liquid, and gas are called *phases of matter*. Going from one phase to another in everyday life is called a *phase transition*. Under conditions of extreme pressure and temperature, phase transitions do not occur. As a result, on Jupiter there is no definite boundary between the planet's gaseous atmosphere and its liquid mantle. The hydrogen gradually gets denser until 1000 km (620 mi) below the cloud tops the pressure is high enough for the hydrogen to be what we would consider a liquid.

In introducing the solar system, we noted that the young planets heated up as they coalesced. After they formed, radioactive elements continued to heat their interiors. On Earth, this heat leaks out of the surface through volcanoes and other vents. Jupiter loses heat through its clouds everywhere because, unlike Earth, it has no surface landmasses to block the heat loss.

While being heated from deep within Jupiter, blobs of liquid hydrogen and helium move upward inside the planet. When these blobs reach the cloud tops, they release their heat and descend back into the interior. (This same process of *convection* drives the motion of Earth's mantle and its tectonic plates, as in liquid simmering on a stove; see Figure 6-11.) Jupiter's rapid, differential rotation draws the convective gases into bands of winds moving eastward and westward at different speeds around the planet.

Astronomers believe that Jupiter's belts and zones result from the combined actions of the planet's con-

vection and rapid differential rotation. The light-colored zones are regions of hotter, rising gas, while the dark belts are regions of cooler, descending gas (Figure 7-7). Intriguingly, high-resolution observations by the passing *Cassini* spacecraft on its way to Saturn revealed numerous white clouds that are too small to be seen from Earth rising in the dark belts of Jupiter. Astronomers are working to understand these newly discovered features.

7-2 Jupiter's interior has four distinct regions

Descending below Jupiter's clouds, we first encounter liquid molecular hydrogen and helium. As in Earth's oceans, the pressure in this fluid increases with depth. The gravitational force created by Jupiter's enormous mass compresses and heats its interior so much that 20,000 km (12,500 mi) below the cloud tops, the pressure is 3 million atm. Below this depth, the pressure is

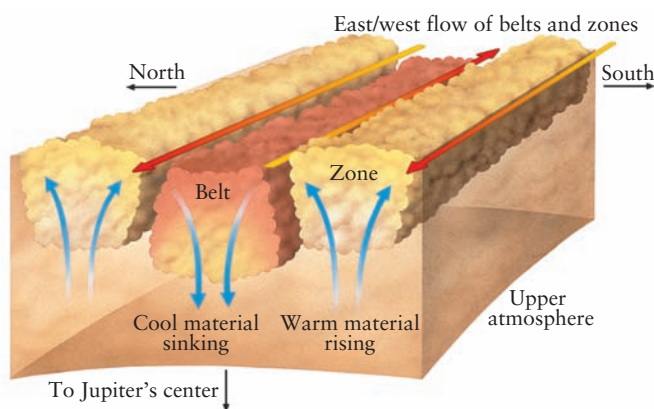


FIGURE 7-7 Jupiter's Belts and Zones The light-colored zones and dark-colored belts in Jupiter's atmosphere are regions of rising and descending gases, respectively. In the zones, gases warmed by heat from Jupiter's interior rise upward and cool, forming high-altitude clouds. In the belts, cooled gases descend and undergo an increase in temperature; the cloud layers seen there are at lower altitudes than in the zones. Jupiter's rapid differential rotation shapes the rising and descending gas into bands of winds parallel to the planet's equator.

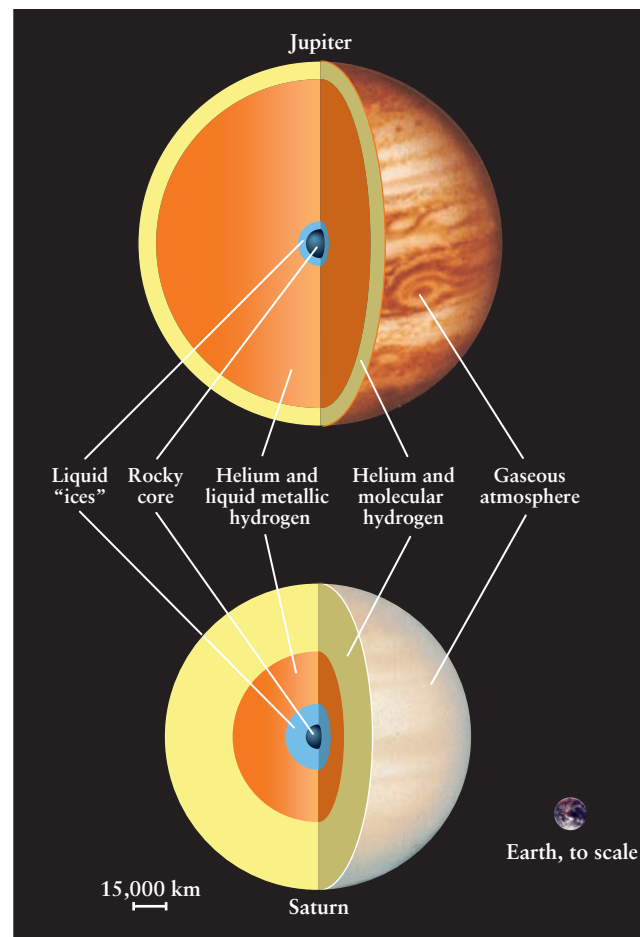
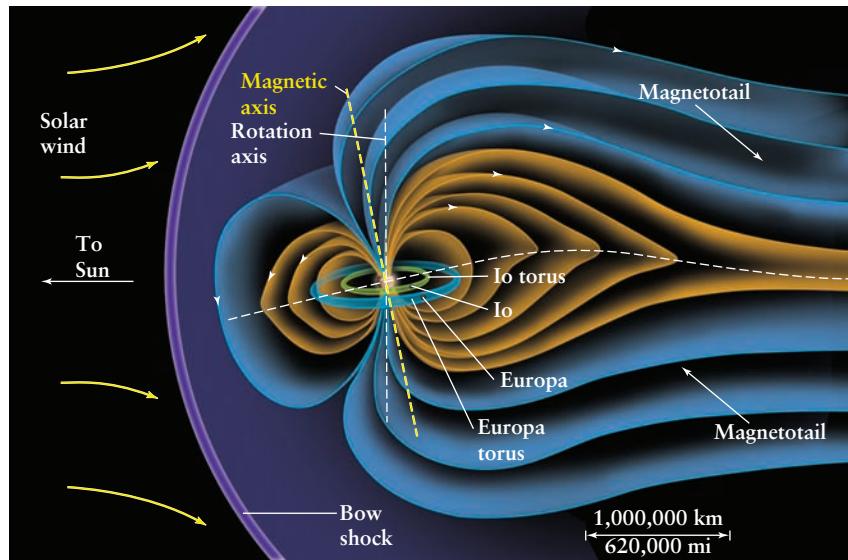


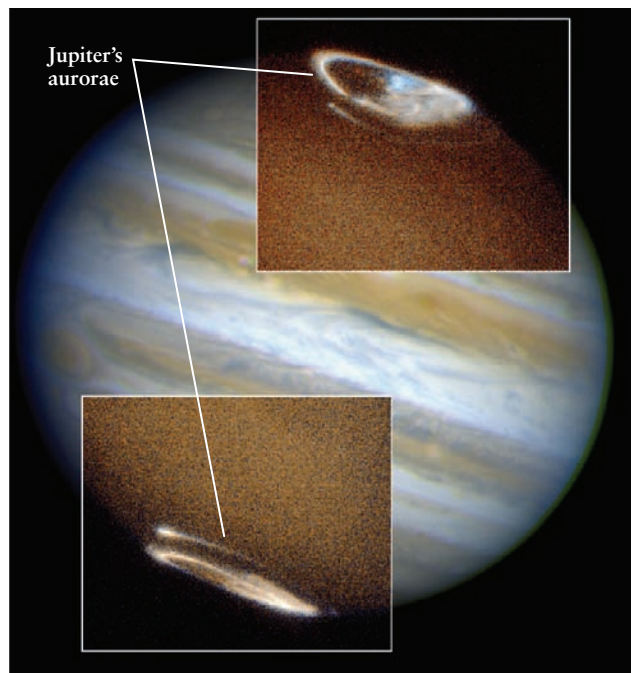
FIGURE 7-8 Cutaways of Jupiter and Saturn The interiors of both Jupiter and Saturn are believed to have four regions: a terrestrial rocky core, a thin shell of liquid water, carbon dioxide, methane, and ammonia (collectively called ices), a liquid metallic hydrogen shell, and a normal liquid hydrogen mantle. Their atmospheres are thin layers above the normal hydrogen, which convects and thereby creates the belts and zones.

high enough to transform hydrogen into **liquid metallic hydrogen** (Figure 7-8). In this state, hydrogen acts like a metal (similar to the copper wiring in a house) in its

ability to conduct electricity and heat. Electric currents that run through this rotating, metallic region of Jupiter generate a powerful planetary magnetic field (Figure 7-9).

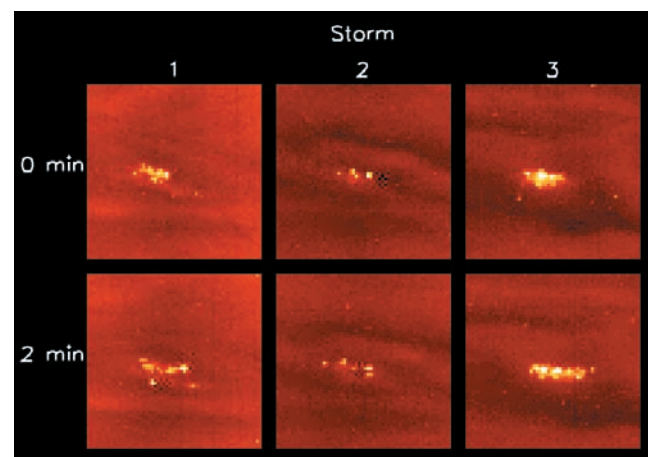


a



b

R I V U X G



c

FIGURE 7-9 Jupiter's Magnetosphere (a) Created by the planet's rotation, the ion-trapping regions of Jupiter's magnetosphere (in orange, analogous to the Van Allen belts) extend into the realm of the Galilean moons. Gases from Io and Europa form tori (doughnut-shaped regions) in the magnetosphere. Some of Io's particles are pulled by the field onto the planet. Pushed outward by the solar wind, the magnetosphere has a "magnetotail" pointing away from the Sun. The magnetotail is often over 500 million km (310 million mi) long and sometimes it reaches all the way to Saturn. (b) High-energy particles,

trapped in Jupiter's magnetosphere, excite gases in the planet's upper atmosphere, causing them to glow as aurorae. The magnetosphere and cloud motion also lead to lightning on Jupiter. (c) Three lightning storms originating in Jupiter's water cloud layer, imaged by the *Galileo* spacecraft. Each image shows changing, multiple lightning flashes. The brightest images are lightning flashes as powerful as 30 million 100-watt lightbulbs shining for 1 second. This is much more powerful than any lightning on Earth. (b: J. Clarke, University of Michigan, and NASA; c: NASA/JPL-Caltech)

At the cloud level of Jupiter, this field is 14 times stronger per square meter than Earth's field is at our planet's surface. Jupiter's magnetic field has a region, like Earth's Van Allen belts, where charged particles are stored (depicted in orange in Figure 7-9a) and its tail (called the *magnetotail*) streams outward as far as Saturn, over 700 million km (435 million mi) away.

For Earth-based astronomers, the first evidence of Jupiter's magnetosphere was a hiss of radio static, which varies cyclically over a period of 9 h 55 min 30 s, the planet's internal rotation rate. In recent years, astronomers have also been able to observe from Earth (and from space) permanent aurorae surrounding Jupiter's poles (Figure 7-9b). These aurorae have their distinctive ring shapes as a result of the glowing gases being trapped by the planet's magnetic field. Beginning with the two *Pioneer* and two *Voyager* spacecraft that journeyed past Jupiter in the 1970s, and continuing with the recent *Galileo* and *Cassini* spacecraft, the awesome dimensions of Jupiter's magnetosphere have been revealed: It is nearly 30 million km (19 million mi) across and envelops the orbits of many of its moons. If Jupiter's magnetosphere were visible from Earth, it would cover an area of the sky 16 times larger than the Moon as seen from here.

Based on the Nice theory of giant planet formation (Section 4-3), it appears that a terrestrial (rock and metal) protoplanet formed out beyond the snow line and attracted hydrogen and helium to form its outer layers. This terrestrial matter is now Jupiter's core. This core is only 4% of Jupiter's mass, which still amounts to nearly 13 times the mass of the entire Earth. Water, carbon dioxide, methane, and ammonia are likely to have existed as ice on Jupiter's terrestrial protoplanet. When astronomers talk in general terms about "ice," they are referring to any or all of these four compounds.

The tremendous crushing weight of the bulk of Jupiter above the core—equal to the mass of 305 Earths—compresses the terrestrial core down to a sphere only 10,000 km (6200 mi) in radius (see Figure 7-8). By comparison, Earth's diameter is 12,756 km (7926 mi). The pressure at Jupiter's very center is calculated to be about 70 million atm, and the temperature there is about 25,000 K, nearly 4 times hotter than the surface of the Sun. Recent calculations using sophisticated computer simulations of Jupiter's interior have discovered that the iron in its core is likely dissolving as it interacts with the surrounding liquid metallic hydrogen. This result is going to be explored in 2016 when the NASA spacecraft *Juno* reaches Jupiter. If the iron is dissolving, then Jupiter's core is shrinking. The interior of Jupiter is summarized in Figure 7-8a.

Interestingly, Jupiter gives off about twice as much energy as it receives from the Sun. Emitting this extra energy causes the planet to cool ever so slightly. As a planet cools, its gravity causes it to contract. Therefore, Jupiter is presently shrinking about a centimeter

per decade. This process of cooling and contracting, called the *Kelvin-Helmholtz mechanism*, was proposed by Lord Kelvin and Hermann von Helmholtz in the nineteenth century.

7-3 Impacts provide probes into Jupiter's atmosphere

On July 7, 1992, a comet nucleus (a clump of rock and ice a few kilometers across, discussed in Chapter 8) passed so close to Jupiter that the planet's gravitational tidal force ripped it into at least 21 pieces. The debris from this comet was first observed in March 1993 by comet hunters Gene and Carolyn Shoemaker and David Levy. (Because it was the ninth comet they had found together, it was named Shoemaker-Levy 9 in their honor.) Shoemaker-Levy 9 was an unusual comet that actually orbited Jupiter rather than just orbiting the Sun. Calculations of the comet's orbit showed that the pieces (Figure 7-10a) would return to strike Jupiter between July 16 and July 22, 1994.

Recall from Chapter 4 that impacts were extremely common in the first 800 million years of the solar system's existence. From more recent times, two chains of impact craters have been discovered on Earth, consistent with pieces of comets having hit our planet within the past 300 million years. However, it is very uncommon for pieces of space debris as large as several kilometers in diameter to collide with planets today. Therefore, the discovery that Shoemaker-Levy 9 would hit Jupiter created great excitement in the astronomical community. Seeing how a planet and a comet respond to such an impact would allow astronomers to deduce information about the planet's atmosphere and interior as well as about the striking body's properties.

The impacts occurred as predicted, with most of Earth's major telescopes—as well as those on several spacecraft—watching closely (Figure 7-10b). At least 20 fragments from Shoemaker-Levy 9 struck Jupiter, and 15 had detectable impact sites. The impacts resulted in fireballs some 10 km (6 mi) in diameter with temperatures of 7500 K, which is hotter than the surface of the Sun. Indeed, the largest comet fragment gave off as much energy as 600 million megatons of TNT, far more than the combined energy that could be released by all the nuclear weapons currently stockpiled on Earth. Impacts were followed by crescent-shaped ejecta that contained a variety of chemical compounds. Ripples or waves spread out from the impact sites through Jupiter's clouds in splotches that lasted for months.

The observations suggested that the pieces of comet did not penetrate very far into Jupiter's upper cloud layer. This fact, in turn, supports the belief that the pieces were not much larger than a kilometer in diameter. The ejecta from each impact included a dark plume that rose high into Jupiter's atmosphere. The darkness was apparently due to carbon compounds that vaporized from the come-

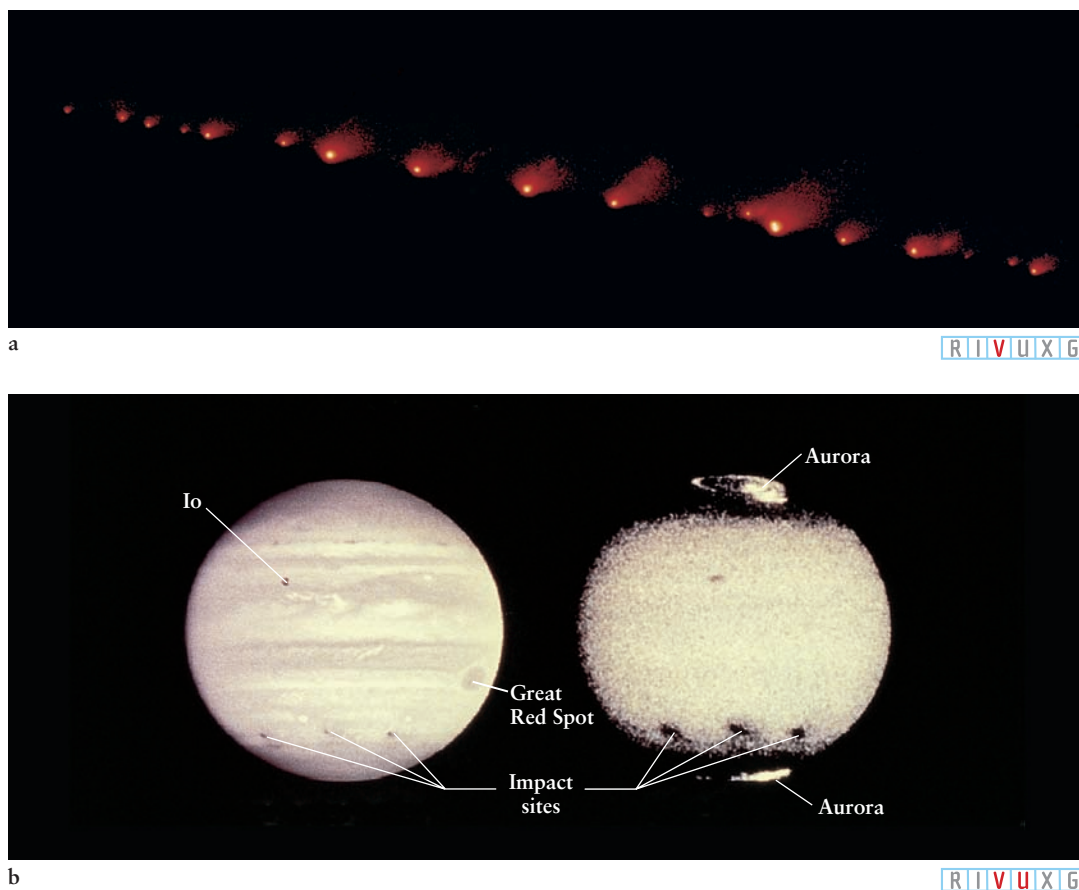


FIGURE 7-10 Comet Shoemaker-Levy 9 and Its Encounter with Jupiter (a) This comet, originally orbiting Jupiter, was torn apart by the planet's gravitational force on July 7, 1992, fracturing into at least 21 pieces. Its returning debris, shown here in May 1994, struck the planet between July 16 and July 22, 1994. (b) Shown here are visible (left) and ultraviolet (right) images of Jupiter taken by the Hubble

Space Telescope after three pieces of Comet Shoemaker-Levy 9 struck the planet. Astronomers had expected white remnants (the color of condensing ammonia or water vapor); the darkness of the impact sites may have come from carbon compounds in the comet debris. Note the aurorae in the ultraviolet image. (a: H. A. Weaver, T. E. Smith, STScI and NASA; b: NASA)

tary bodies. Also detected from the comet collisions were water, sulfur compounds, silicon, magnesium, and iron.

On July 19, 2009, an amateur astronomer discovered a new dark patch near Jupiter's south pole. Further observations revealed that this feature was created by another impact that had occurred about 4 hours before it was actually observed. Unlike the impact of 1994, this collision occurred when a small asteroid, about 0.5 km (0.3 mi) in diameter and composed of rock and metal, struck Jupiter. Astronomers were able to distinguish between the two causes (comet versus asteroid) by noting that gases and dust drifting along with the pieces of comet that struck Jupiter created halos around the impact sites, while the asteroid, lacking the surrounding gas and dust as it slammed into Jupiter, had no halo. Based on the 1994 and 2009 impacts, it appears that Jupiter is struck by pieces of debris 0.5 to 1 km (0.3 to 0.6 mi) in diameter every 10 to 15 years.

JUPITER'S MOONS AND RINGS

Jupiter hosts at least 67 moons. The four largest—Io, Europa, Ganymede, and Callisto—may have formed at the same time as Jupiter from debris orbiting the planet, just as the planets formed from debris orbiting the Sun (see Chapter 4). Jupiter's other moons are probably captured planetesimals and smaller pieces of space debris.

Galileo was the first person to observe the four largest moons, in 1610, seen through his meager telescope as pinpoints of light. He called them the *Medicean stars* to attract the attention of the Medicis, rulers of Florence and wealthy patrons of the arts and sciences. To Galileo, the moons provided evidence supporting the then-controversial Copernican cosmology; at that time, Western theologians asserted that all cosmic bodies orbited Earth. The fact that the Medicean stars orbited Jupiter raised grave concerns in some circles.

Focus Question 7-2

Why did the comet impacts of 1994 not create craters on Jupiter?

To the modern astronomer, these moons are four extraordinary worlds, different both from the rocky terrestrial planets and from hydrogen-rich Jupiter. Now called collectively the **Galilean moons** or **Galilean satellites**, they are named after the mythical lovers and companions of the Greek god Zeus (called Jupiter by the

Romans). From the innermost moon outward, they are Io, Europa, Ganymede, and Callisto. All four Galilean moons orbit in the same direction and in the same plane. These four worlds were photographed extensively by the *Voyager 1*, *Voyager 2*, and *Cassini* flybys and by the *Galileo* spacecraft that orbited Jupiter (Figure 7-11). The

THE GALILEAN MOONS, MERCURY, AND EARTH’S MOON: Vital Statistics

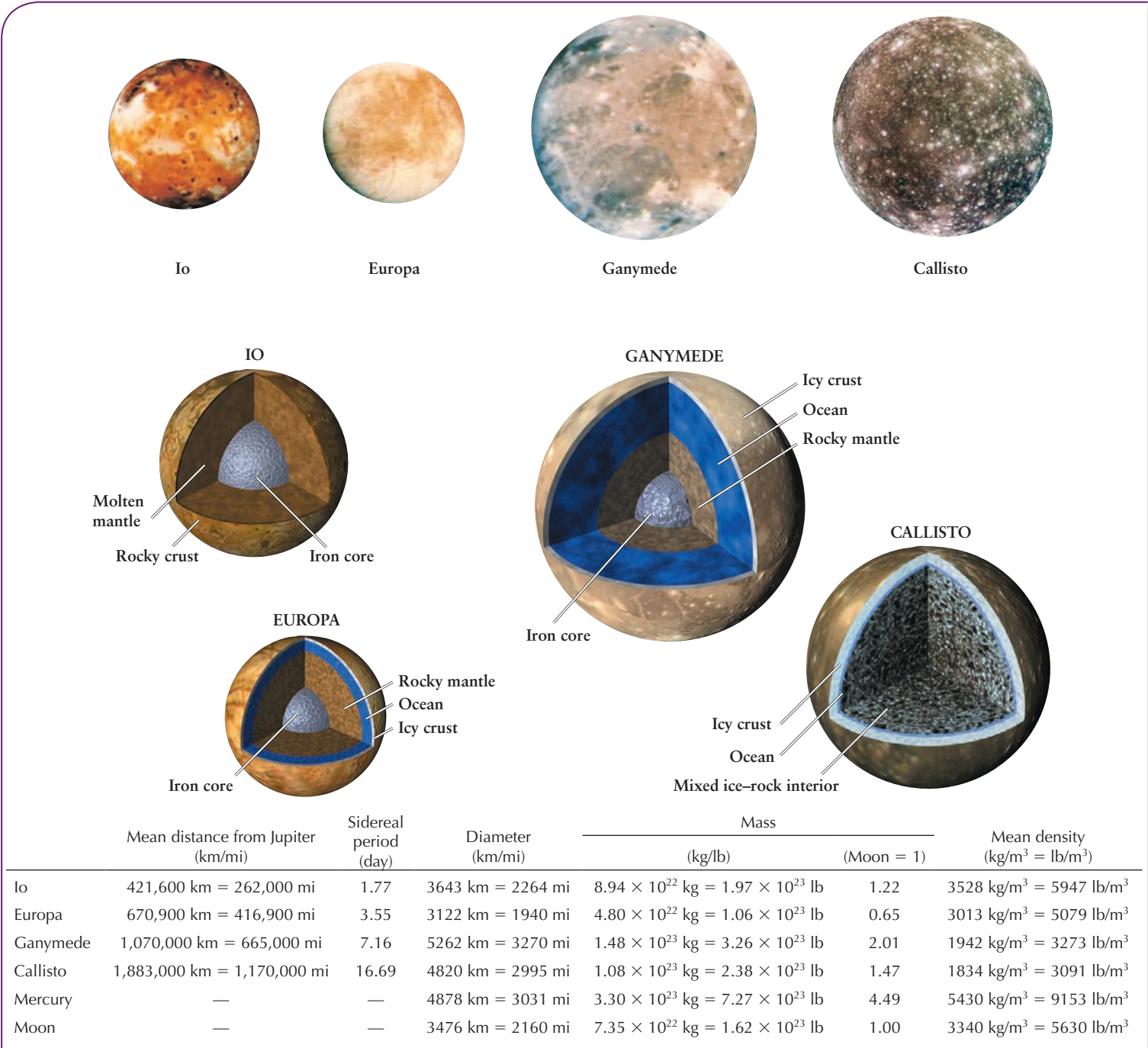


FIGURE 7-11 The Galilean Satellites The four Galilean satellites are shown here to the same scale. Io and Europa have diameters and densities comparable to our Moon and are composed primarily of rocky material. Ganymede and Callisto are roughly as big as Mercury, but their low average densities indicate that each contains a thick layer of

water and ice. The cross-sectional diagrams of the interiors of the four Galilean moons show the probable internal structures of the moons, based on their average densities and on information from the *Galileo* mission. (a–c and e–h: NASA and NASA/JPL; d: NASA/JPL/DLR)

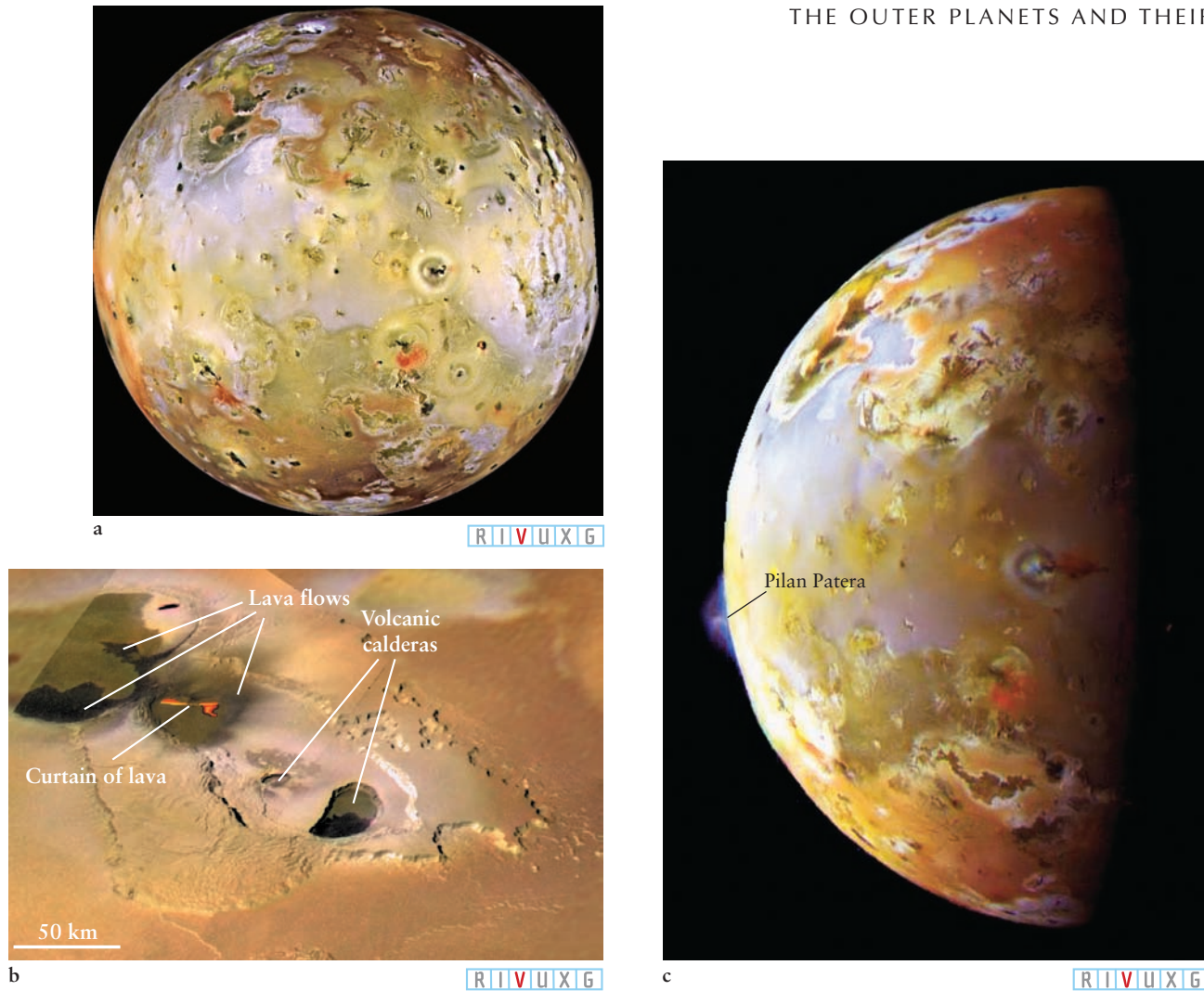


FIGURE 7-12 Io (a) This striking *Galileo* image of Io shows a variety of surface features including volcanoes, snow fields, and solidified lava flows. The range of colors results from surface deposits of sulfur ejected from Io's numerous volcanoes. Plumes from the volcano Prometheus rise up 100 km (62 mi). Prometheus has been active in every image taken of Io since

the *Voyager* flybys of 1979. (b) Photographed in 1999 and then 2000 (shown here), the ongoing lava flow from this volcanic eruption at Tvashtar Patera has considerably altered this region of Io's surface. (c) *Galileo* image of an eruption of Pilon Patera on Io. (a: NASA/JPL/University of Arizona; b: NASA/JPL/University of Arizona; c: *Galileo* Project, JPL, NASA)

two inner Galilean satellites, Io and Europa, are about the same size as our Moon. The two outer satellites, Ganymede and Callisto, are roughly the size of Mercury. Figure 7-11 presents comparative information about these six bodies.

7-4 Io's surface is sculpted by volcanic activity

Sulfurous Io is among the most exotic moons in our solar system (Figure 7-12a). At first glance, this moon's density of 3570 kg/m^3 (6020 lb/yd^3) seems to place its chemical composition between that of the terrestrial and Jovian planets. However, its density has that value in large measure because its mass is so small that it cannot compress its interior nearly as much as can Earth or more massive planets. Allowing for its lower compression, astronomers calculate that Io is mostly rock and iron, like Earth, rather than being composed of lighter elements, as is Jupiter.

Insight Into Science

What's in a Name? Following up on the preconceptions that common words create (see Insight Into Science: Imagine the Moon in Chapter 6), objects with familiar names often have different characteristics than we expect. We tend to envision "moons" as spherical, inert, airless, lifeless, dry places similar to our Moon. However, as we will see throughout this chapter, starting with Io, some moons have active volcanoes, atmospheres, and vast underground liquid water oceans. We will also see that most moons are anything but spherical.

Io zooms through its orbit of Jupiter once every 1.8 days. Like our Moon, it is in synchronous rotation with its planet. Images of Io reveal giant plumes emitted by about 400 active volcanoes (Figure 7-12c) and

lava geysers (analogous to the water geyser Old Faithful on Earth). Most of this ejected material falls back onto Io's surface; the rest is moving fast enough to escape into space. Besides the violent ejections of lava, Io also has volcanoes out of which lava just pours rather than spurts. Observations indicate that Io's volcanoes emit 9×10^{15} kg (10 trillion tons) of lava each year in plumes up to 500 km (310 mi) high. This is 100 times the amount of lava that is emitted by all of Earth's volcanoes, which is enough to resurface Io to a depth of 1 m (1.1 yd) each century (Figure 7-12b). One eruption in 2002 occurred over an area the size of London, some 1600 km² (620 mi²). The volcanoes emit basaltic lava flows rich in magnesium and iron. Io's volcanoes are named after gods and goddesses associated with fire in Greek, Norse, Hawaiian, and other mythologies. Io also has numerous black "dots" on its surface that apparently are dormant volcanic vents. Old lava flows radiate from many of these locations, which are typically 10 to 50 km (6.2 to 31 mi) in diameter and cover 5% of Io's surface.

Just before their discovery, the existence of active volcanoes on Io was predicted from analysis of the gravitational forces to which that moon is subjected. As it rapidly orbits Jupiter, Io repeatedly passes between Jupiter and one or another of the remaining Galilean satellites. These moons pull on Io, causing it to slightly change its

distance from Jupiter. This change in distance creates a change in the tidal forces (see Section 6-8) acting on it from the planet. As the distance between Io and Jupiter varies, the resulting tidal stresses alternately squeeze and flex the

moon. This ongoing motion of Io's interior generates heat through friction, creating as much energy inside Io as the detonation of 2400 tons of TNT every second. As a result, Io has a layer of magma (molten underground material that becomes lava on the surface) some 30–50 km (20–30 mi) below the surface. Gas and magma from it eventually make their way to the moon's surface, where they are ejected through volcanoes and other vents.

Satellite instruments have identified sulfur and sulfur dioxide in the material erupting from Io's volcanoes. Sulfur is normally bright yellow. If heated and suddenly cooled, however, it forms molecules that assume a range of colors, from orange and red to black, which accounts

for Io's tremendous range of colors (see Figure 7-12). Sulfur dioxide (SO₂) is an acrid gas commonly discharged from volcanic vents here on Earth and, apparently, on Venus. When eruptions on Io release this gas into the cold

vacuum of space, the gas crystallizes into white flakes, which fall onto the surface and account for the moon's whitish deposits.

The *Galileo* spacecraft detected an atmosphere around Io. Composed of oxygen, sulfur, and sulfur diox-

ide, it is only one-billionth as dense as the air we breathe. Io's atmosphere can sometimes be seen to glow blue, red, or green, depending on the gases involved. Gases ejected from Io's volcanoes have also been observed extending out into space and forming a doughnut-shaped region around Jupiter called the *Io torus*, about which we will say more shortly.

7-5 Europa harbors liquid water below its surface

Images of Europa's ice and rock surface from *Voyager 2* and the *Galileo* spacecraft (Figure 7-13) reveal that the interior of Jupiter's second-closest Galilean moon contains liquid water (see Figure 7-11). Europa orbits Jupiter every 3½ days and, like Io, is in synchronous rotation. The changing gravitational tug of Io creates stress inside Europa similar to the magma-generating distortion that Io undergoes. These stresses apparently cause numerous cracks or fractures seen on Europa's surface (Figure 7-14a). Furthermore, astronomers think this stress creates the heat inside Europa that is keeping water in a liquid state just a few kilometers below the moon's ice and rock surface.

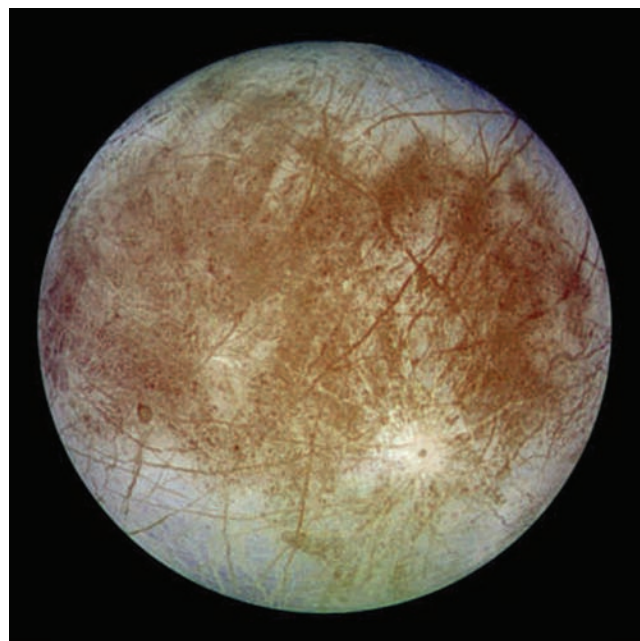
Beside the cracks, numerous other features on Europa support the idea that it has liquid water inside. Just as in the Arctic region on Earth, the movement of surface features on Europa creates ice floes, along with swirls, strips, and ridges, driven by circulating water underneath the moon's surface (Figure 7-13, 7-14b). Also indicative

Focus Question 7-3

What creates most of the heat inside Io that causes it to have volcanoes and geysers?

Focus Question 7-4

What region on Earth does the surface of Europa most resemble?



RIVUXG

FIGURE 7-13 Europa Imaged by the *Galileo* spacecraft, Europa's ice surface is covered by numerous streaks and cracks that give the satellite a fractured appearance. The streaks are typically 20 to 40 km (12 to 25 mi) wide. (NASA/JPL/DLR)

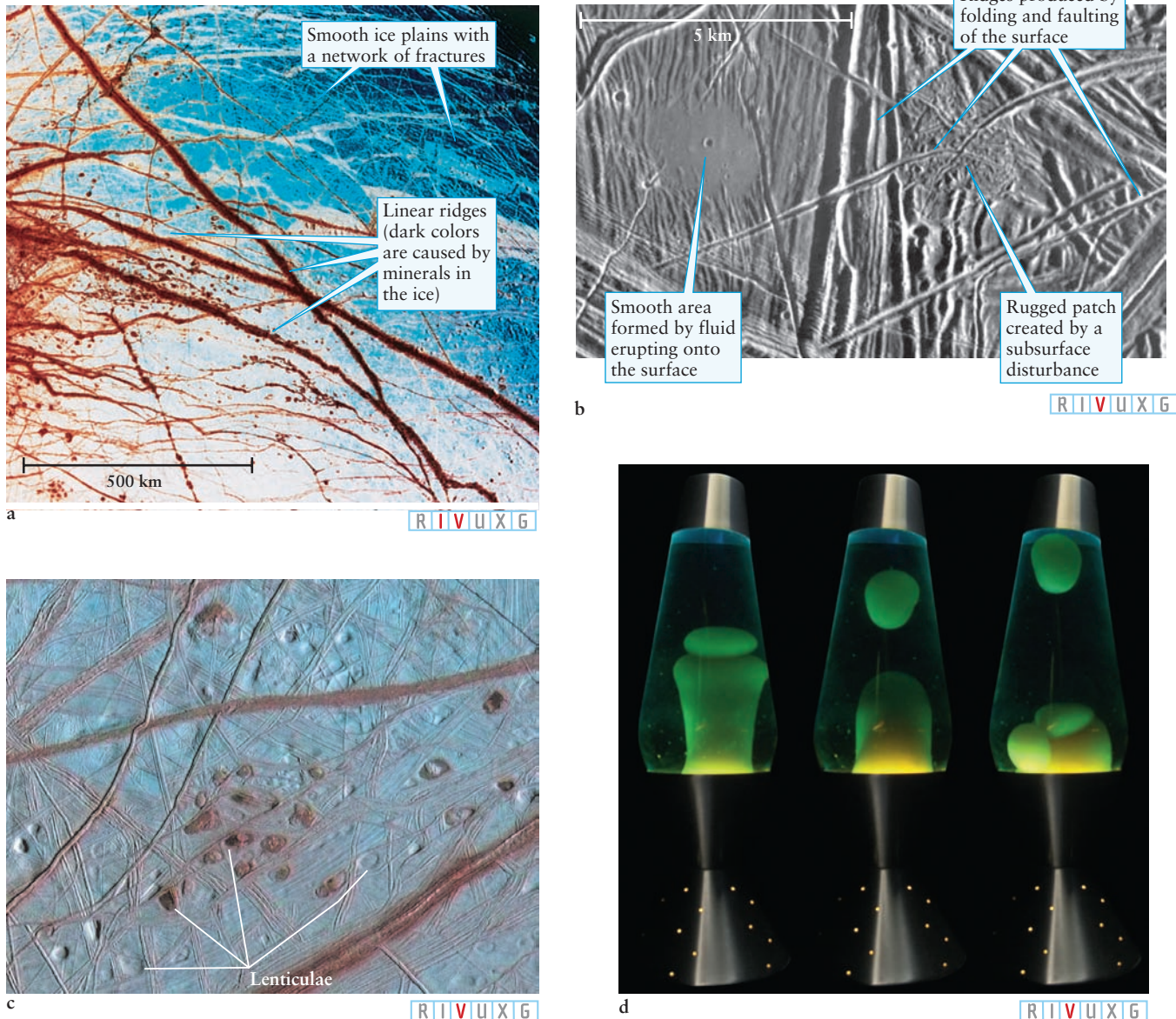


FIGURE 7-14 Surface Features on Europa (a) This false-color *Galileo* image of Europa combining visible and infrared observations shows smooth plains of ice, mineral ridges deposited by upwelling water, and numerous fractures believed to be caused by tidal stresses. (b) This region of Europa's surface shows the jumbled, stressed features common to the surface, as well as direct indications of liquid water activity underground. (c) Lenticulae attributed to rising warmed ice and debris travel

of liquid water is the moon's reddish color. The coloring may be due to salt deposits left after liquid water rose to the surface and evaporated (Figures 7-13, 7-14a, c). *Galileo* spacecraft images suggest that some of Europa's features have moved within the past few million years and perhaps are still in motion. Indeed, the chaotic surface revealed by *Galileo* is interpreted as having formed as a result of water volcanism on Europa, strengthening the belief that this moon still has a liquid water layer.

up from the moon's interior by convection, arriving at and then leaking out at the surface. The white domes are likely to be rising material that has not yet reached the surface. (d) A lava lamp in which warmed material rises through cooler liquid. The rising material in Europa is analogous to the rising motion of the blobs in a lava lamp, except that on Europa, the motion is through ice. (a and b: NASA/JPL; c: NASA/JPL/University of Arizona/University of Colorado; d: Bianca Moscatelli)

Replenishment of the surface by tectonic plate motion would explain why only a few small impact craters have survived. In 2013, the Hubble Space Telescope observed water vapor above this moon's south polar region, confirming the existence of Europa's liquid water.

Galileo also photographed red and white domes called *lenticulae* on Europa (Figure 7-14c) that geologists think are rising warmed ice mixed with other material, possibly including organic matter. These lenticulae are

typically 100 m (109 yd) high and 10 km (6 mi) wide. The rising material behaves like the blobs in lava lamps (Figure 7-14d), although the lenticulae are calculated to take 100,000 years to reach the surface from the liquid ocean inside the moon.

Europa's average density of 2970 kg/m^3 (5000 lb/yd^3) is slightly less than on Io. A quarter of its mass may be water. Europa also has a metallic core of much higher density than Io and a weak magnetic field. In 1995, astronomers discovered an extremely thin atmosphere containing molecular oxygen surrounding Europa. This gas is about 10^{-11} times as dense as the air we breathe. The oxygen may come from water molecules on the moon's surface broken up by ultraviolet radiation from the Sun.

The underground ocean of Europa is especially interesting to scientists because virtually all liquid water locations on Earth support life, and life may well have evolved in Europa's oceans. If it exists, and if it evolved differently from life on Earth, then life-forms from Earth would likely contaminate it, potentially wiping it out. This process would be analogous to the deadly impact of diseases on Earth, such as the introduction of smallpox to the Americas from Europe. To prevent the expired spacecraft *Galileo* from contaminating any life that may exist on Europa, NASA sent that spacecraft into Jupiter's thick atmosphere where the machine was vaporized.

7-6 Ganymede is larger than Mercury

Ganymede is the largest satellite in the solar system (Figure 7-15). Its diameter is greater than Mercury's, but its density of 1940 kg/m^3 (3270 lb/yd^3) is much less than that of Mercury. This moon also has a permanent magnetic field that is twice as strong as Mercury's field. Ganymede orbits Jupiter in synchronous rotation once every 7.2 days. Like its neighbor Europa, Ganymede has an iron-rich core, a rocky mantle, an underground liquid water ocean, a thin atmosphere, and a covering of dirty ice (see Figure 7-11).

The existence of an underground ocean is implied by the discovery of a second magnetic field around Ganymede that continually varies. This field has a different origin than the permanent one just mentioned; the second field is generated by Jupiter's magnetic field. As Ganymede orbits Jupiter, the planet's powerful magnetic field

creates an electric current inside the moon, which, in turn, creates Ganymede's varying magnetic field. The field varies as the moon rotates and as it revolves around Jupiter. The best explanation of why current flows inside Ganymede is that liquid saltwater exists there, and saltwater is a good conductor of electricity. This implies, of course, the presence of a liquid ocean. Furthermore, salts have been observed on Ganymede's surface. They were apparently carried upward and

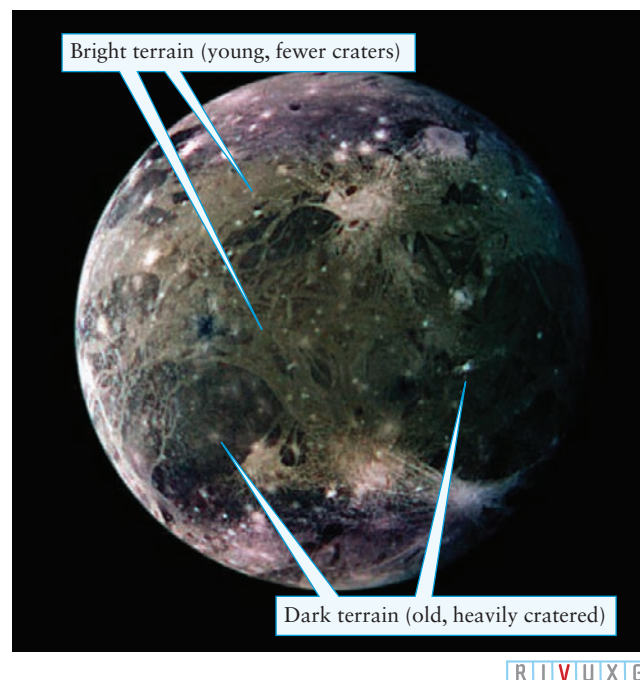


FIGURE 7-15 Ganymede This side of Ganymede is dominated by the huge, dark, circular region called Galileo Regio, which is the largest remnant of Ganymede's ancient crust. Darker areas of the moon are older; lighter areas are younger, tectonically deformed regions. Large impacts create white craters, filled in by ice from below the surface. (Galileo Project, DLR, JPL, NASA)

deposited there as water leaked out and froze. As on Europa, liquid water may imply the presence of life.

Like our Moon, Ganymede has two very different kinds of terrain (Figure 7-15 and Figure 7-16). Dark, polygon-shaped regions are its oldest surface features, as judged by their numerous craters. Light-colored, heavily grooved terrain is found between the dark, angular islands. These lighter regions are much less cratered and therefore younger. Ganymede's grooved terrain consists of parallel mountain ridges up to 1 km (0.6 mi) high and spaced 10 to 15 km (6 to 9 mi) apart. These features suggest that the process of plate tectonics may have dominated Ganymede's early history. But, unlike Europa where tectonic activity still occurs today, tectonics on Ganymede ceased 3 billion years ago as the satellite's crust froze solid.

Another mechanism that is likely to have helped create Ganymede's large-scale features originated in a bizarre property of water. Unlike most liquids, which shrink upon solidifying, water expands when it freezes. Seeping up through cracks in Ganymede's original crust and solidifying, water may have forced apart fragments of that crust. This process could have produced jagged, dark islands of old crust separated by bands of younger, light-colored, heavily grooved ice.

Focus Question 7-5

Which planet is smaller than Ganymede?

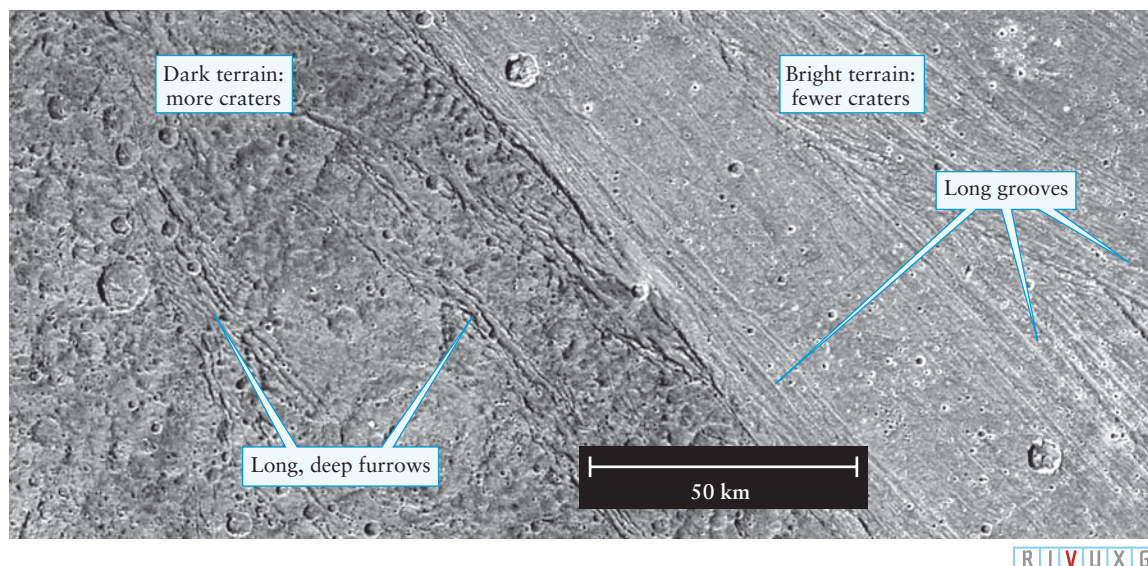


FIGURE 7-16 Two Surfaces of Ganymede The older, rougher, more heavily cratered parts of Ganymede are the dark terrain. These regions are surrounded by younger, smoother,

less cratered bright terrain. The parallel ridges suggest that the bright terrain has been crafted by tectonic processes. (NASA/JPL)

Ganymede has a thin atmosphere known to contain oxygen and possibly other gases. As it passes through Jupiter's magnetic field, that oxygen is bombarded by electrons trapped by the field. These collisions cause the gas in Ganymede's atmosphere to give off radiation, creating auroras, which were first observed in 1998.

7-7 Callisto bears the scars of a huge asteroid impact

Callisto is Jupiter's outermost Galilean moon. It orbits Jupiter in 16.7 days and, like the other Galilean moons, its rotation is synchronous. Callisto is 91% as big and 96% as dense as Ganymede and has a thin atmosphere of hydrogen and carbon dioxide. Like Ganymede, Callisto apparently harbors a substantial liquid water ocean. As with Ganymede, this ocean's presence is inferred by Callisto's changing magnetic field. The heat that keeps the ocean liquid likely comes from energy released by radioactive decay inside the moon.

Although numerous large impact craters are scattered over Callisto's dark, ancient, icy crust, it has very few craters smaller than 100 m (109 yd) across (Figure 7-17). Astronomers speculate that the smaller craters have disintegrated. Unlike Ganymede and Europa, Callisto has no younger, grooved terrain. The absence of grooved terrain suggests that tectonic activity never began there, probably because the satellite simply froze too rapidly. It is bitterly cold on Callisto's surface. *Voyager* instruments measured a noontime temperature of 155 K (−180°F), and the nighttime temperature plunges to 80 K (−315°F).

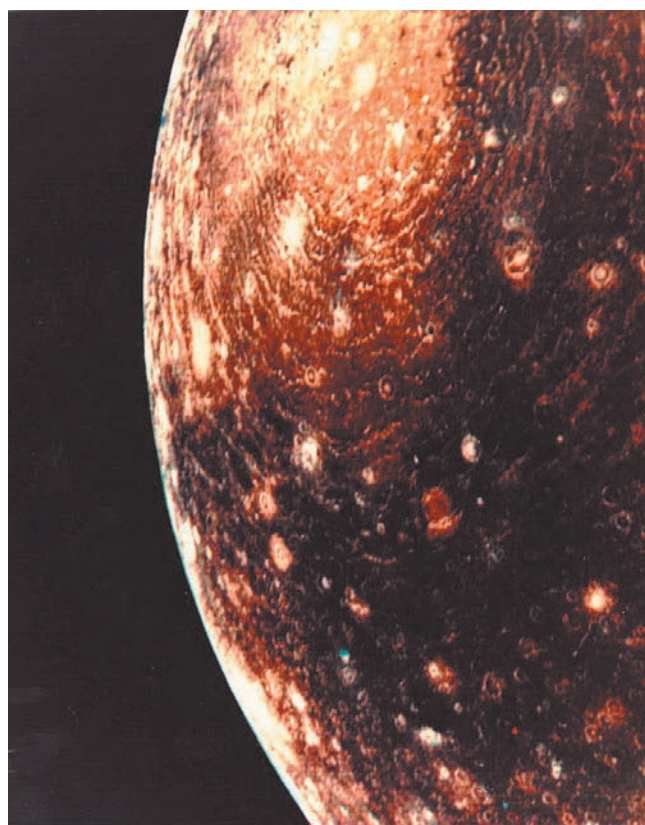
Like Mercury, our Moon, and other bodies we will encounter shortly, Callisto carries the cold, hard evidence of what happens when one astronomical body strikes another. *Voyager 1* photographed a huge impact basin named Valhalla on Callisto (see Figure 7-17a). An asteroid-sized object produced Valhalla Basin, which is located on Callisto's Jupiter-facing hemisphere. Like throwing a rock into a calm lake, ripples ran out from the impact site along Callisto's surface, cracking the surface and freezing into place. The largest remnant rings surrounding the impact crater have diameters of 3000 km (1900 mi). In 2001, the *Galileo* spacecraft revealed spires 80–100 m (87 to 109 yd) high on Callisto (Figure 7-17b). These spires are also believed to have been created by an impact, perhaps the same one. One feature that is missing from Callisto is the jumbled terrain on the side opposite of Valhalla Basin, as is seen opposite of large impact sites on Mercury and our Moon. The smoothness of Callisto can be explained by a liquid water interior, which would dampen the impact shock and thereby prevent the opposite side from becoming disturbed. This theory, of course, supports the idea that Callisto has liquid water inside. The probable interiors of the Galilean moons are shown in Figure 7-11.

Focus Question 7-6

What features on our Moon and on Mercury are similar to Callisto's Valhalla impact basin?

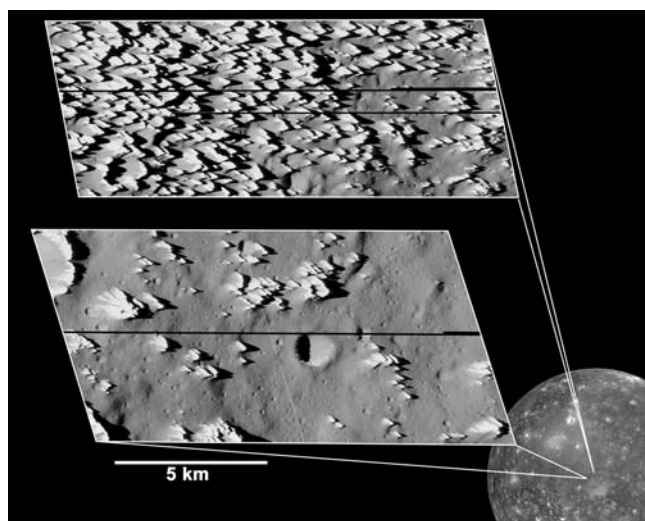
7-8 Other debris orbits Jupiter as smaller moons and ringlets

Along with the four Galilean moons, Jupiter has at least 63 other moons, a set of tenuous *ringlets*, and



a

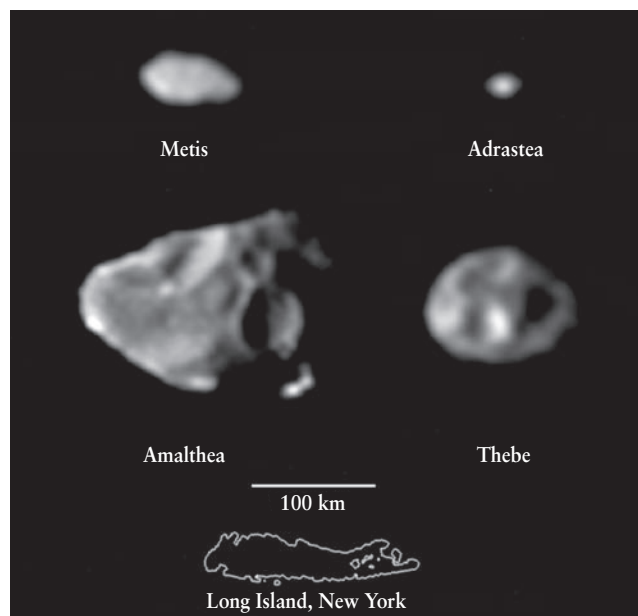
R I V U X G



b

R I V U X G

FIGURE 7-17 Callisto The outermost Galilean satellite is almost exactly the same size as Mercury. Numerous craters pockmark Callisto's icy surface. **(a)** The series of faint, concentric rings that covers much of this image is the result of a huge impact that created the impact basin Valhalla. Valhalla dominates the Jupiter-facing hemisphere of this frozen, geologically inactive world. **(b)** The two insets in this *Galileo* mission image show spires that contain both ice and some dark material. The spires were probably thrown upward as the result of an impact. The spires erode as dark material in them absorbs heat from the Sun. (a: Courtesy of NASA/JPL; b: NASA/JPL/Arizona State University)



R I V U X G

FIGURE 7-18 Irregularly Shaped Inner Moons The four known inner moons of Jupiter are significantly different from the Galilean satellites. These bodies are roughly oval-shaped. Although craters have not yet been resolved on Adrastea and Metis, their irregular shapes strongly suggest that they are cratered. All four moons are named for characters in mythology relating to Jupiter (Zeus, in Greek mythology). (NASA/JPL, Cornell University)

two doughnut-shaped tori of electrically charged gas particles. The non-Galilean moons are all irregular in shape and less than 275 km (171 mi) in diameter. Four of these moons (**Figure 7-18**) are inside Io's orbit; all of the other known moons are outside Callisto's orbit. The inner moon Amalthea (see **Figure 7-18**) is red-colored, has about the same density as water, and is apparently made of pieces of rock and ice barely held together by the moon's own gravity. The Galilean moons, along with the smaller moons closer to Jupiter and six of the outer moons, orbit in the same direction that Jupiter rotates (**prograde orbits**). The remaining outer moons revolve in the opposite direction (**retrograde orbits**). The outer ones appear to be individual, captured planetesimals, whereas the inner ones are probably smaller pieces broken off a single larger body.

4 Jupiter is the first of four planets with rings. Its ring system is composed of several parts (**Figure 7-19a**). Debris knocked off the moons Adrastea and Metis is responsible for the *main* ring. The outer edge of this ring has three well-defined **ringlets** (**Figure 7-19b**) that are pieces of debris ranging in size from gravel to small boulders. The rest of the ring, extending about 10,000 km (6200 mi) inward toward Jupiter, is composed of dust-sized particles. This ring is less than 50 km (30 mi)

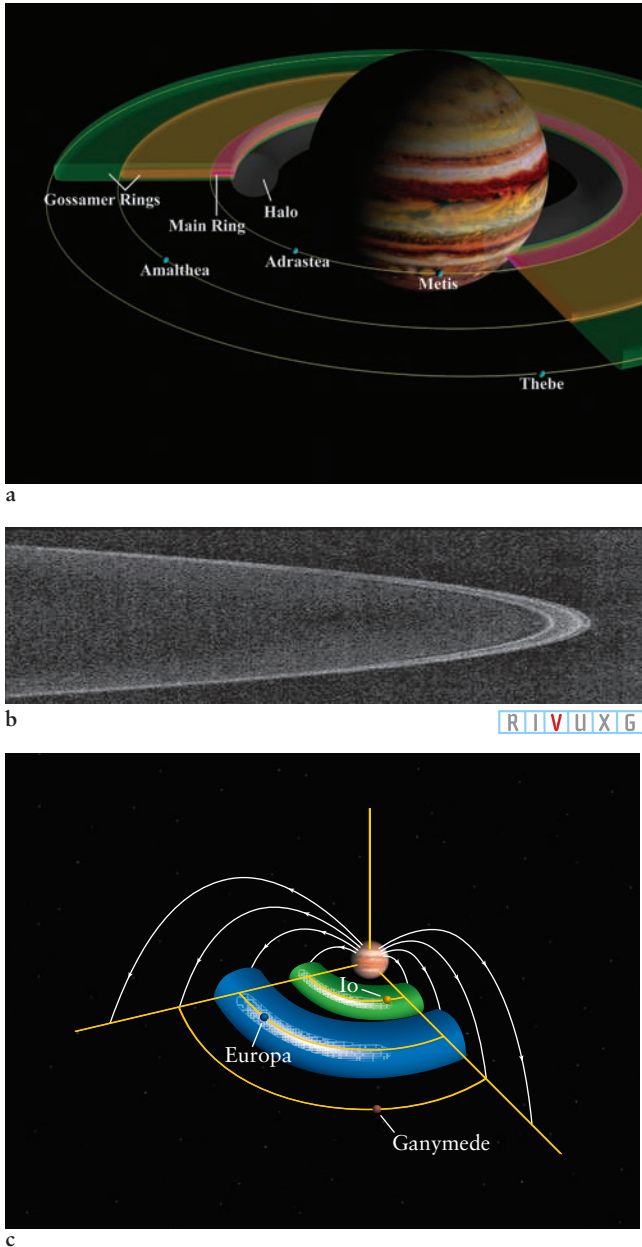


FIGURE 7-19 Jupiter's Ring and Torus (a) Cutaway diagram of Jupiter's rings, which are generated from debris blasted off the inner moons Adrastea, Metis, Amalthea, and Thebe. (b) A portion of Jupiter's faint ring system, photographed by the *New Horizons* spacecraft heading to Pluto. The outer three bright rings are composed of pebble- to rock-sized fragments. The rest is mostly dust. The brightest portion of the ring is about 6000 km (3700 mi) wide. While the outer edge of the ring is sharply defined, the inner edge is somewhat fuzzy. A tenuous sheet of material extends from the ring's inner edge all the way down to the planet's cloud tops. (c) Quarter images of Io's and Europa's torii (also called *plasma tori* because the gas particles in them are charged, that is, the gases are plasmas). Some of Jupiter's magnetic field lines are also drawn in. Plasma from torii flow inward along these field lines toward Jupiter. (a: NASA/JPL/Cornell University; b: NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute; c: NASA/JPL/Johns Hopkins University Applied Physics Laboratory)

thick, which is exceptionally thin compared to its width. Apparently some of the smallest particles leak inward, creating the halo ring (see Figure 7-19a).

Debris ejected from the moons Amalthea and Thebe creates thicker rings of very tiny dust particles called the *gossamer rings*. While some of the ring material is continually being ejected by collisions between particles and by radiation from the Sun, impacts freeing regolith from the moons are continually replenishing all the rings.

At least two doughnut-shaped regions (or tori) of electrically charged gas particles, called *plasmas*, orbit Jupiter. One is in the same orbit as Io (Figure 7-19c). The Io torus consists of sulfur and oxygen ions (charged atoms) along with free electrons. These particles ejected by Io's geysers are held in orbit by Jupiter's strong magnetic field. Guided by the field, some of this matter spirals in toward Jupiter, thereby creating the aurora seen there (see Figure 7-9b) and perhaps contributing also to Jupiter's distinct coloration.

The other torus is in Europa's orbit (Figure 7-19c) and consists of hydrogen and oxygen ions and electrons created from water molecules kicked off Europa's surface by radiation from Jupiter. The mass of this ring of gas is calculated to be only about 5.4×10^4 kg (6×10^4 tons).

Focus Question 7-7

What other objects that we have already studied in the solar system are similar to the smaller moons of Jupiter?

SATURN

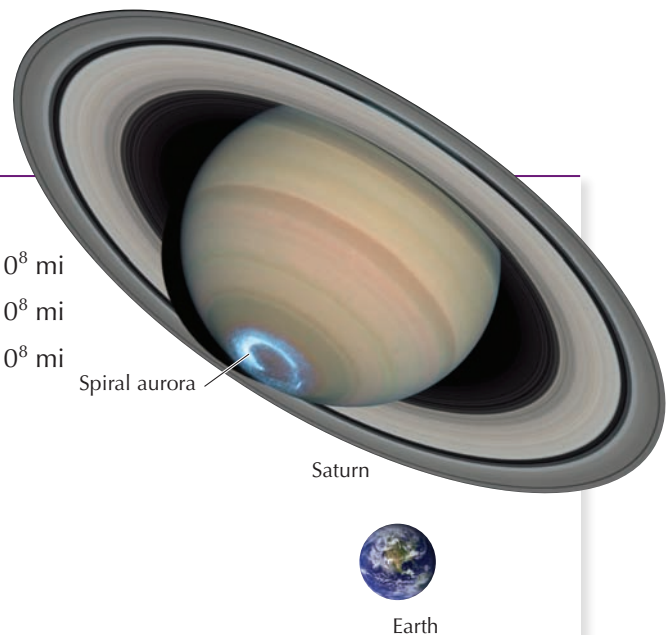
Saturn with its ethereal rings presents the most spectacular image of the planets (Figure 7-20). Giant Saturn has 95 times as much mass as Earth, making it second in mass and size to Jupiter. Like Jupiter, Saturn has a thick, active atmosphere composed predominantly of hydrogen. Indeed, the percentage of hydrogen is higher in Saturn's atmosphere than in Jupiter's atmosphere. Saturn also has a strong magnetic field. Ultraviolet images from the Hubble Space Telescope reveal aurorae around Saturn that appear in spirals (see Figure 7-20), while Jupiter's aurorae appear in rings (see Figure 7-9b).

7-9 Saturn's atmosphere, surface, and interior are similar to Jupiter

Partly obscured by the thick, hazy atmosphere above them, Saturn's clouds lack the colorful contrast visible on Jupiter. Nevertheless, photographs do show faint stripes in Saturn's atmosphere (Figure 7-20), similar to Jupiter's belts and zones. Their existence indicates that Saturn also has internal heat that transports the cloud gases by convection. Changing features in the atmosphere show that Saturn, too, has differential rotation—ranging from about 10 hours and 14 minutes at the

SATURN: Vital Statistics

Planet symbol:	♄
Average distance from the Sun:	9.57 AU = 1.43×10^9 km = 8.89×10^8 mi
Maximum distance from the Sun:	10.1 AU = 1.51×10^9 km = 9.38×10^8 mi
Minimum distance from the Sun:	9.06 AU = 1.36×10^9 km = 8.45×10^8 mi
Orbital eccentricity:	0.056
Average orbital speed:	9.64 km/s = 6.0 mi/s
Orbital period:	29.4 years
Rotation period:	10 h 13 m 59 s (equatorial) 10 h 39 m 25 s (internal)
Inclination of equator to orbit:	26.7°
Inclination of orbit to ecliptic:	2.48°
Diameter:	120,536 km = 74,900 mi = 9.45 Earth diameters (equatorial) 108,680 km = 67,530 mi = 8.52 Earth diameters (polar)
Mass:	5.69×10^{26} kg = 95.2 Earth masses
Average density:	687 kg/m ³ = 1160 lb/yd ³
Escape speed:	35.5 km/s = 22.0 mi/s
Surface gravity (Earth = 1):	0.92
Albedo:	0.46
Average temperature at cloud tops:	−180°C = −2292°F = 93 K
Atmospheric composition (by number of molecules):	96.3% hydrogen (H ₂), 3.3% helium (He), 0.4% methane (CH ₄), ammonia (NH ₃), water vapor (H ₂ O), and other gases



R I V U X G

FIGURE 7-20 Saturn and Its Vital Statistics Combined visible and ultraviolet images from the Hubble Space Telescope reveal spiral arcs of auroras near Saturn’s south pole. The inset image of Earth shows its size

relative to Saturn. Note that there is much less contrast between Saturn’s clouds than those of Jupiter. (NASA, ESA, J. Clarke [Boston University], and Z. Levay [STScI]; inset: NASA)

equator to 10 hours and 40 minutes at high latitudes. As on Jupiter, some of the belts and zones move eastward, while others move westward. In 2006, the *Cassini* spacecraft revealed a hexagonal boundary surrounding Saturn’s north pole (Figure 7-21a). While belts, zones, and large-scale storms (Figure 7-21b) flow around it, this feature is nearly stationary. The mechanism that creates and maintains the boundary is still under investigation.

Although Saturn lacks a long-lived spot like Jupiter’s Great Red Spot, it does have hurricane-like storms. Every 30 years or so (which is one Saturnian year), the planet

Focus Question 7-8
What two effects cause Saturn’s belt and zone system?

experiences a powerful white storm. Lesser storms occur more randomly. In 2004, a pair of storms was observed to merge (Figure 7-22). One of the most powerful storms seen on Saturn in decades occurred throughout most of 2011. It eventually encircled a wide swath of Saturn’s northern hemisphere, with its head eventually overtaking its tail (Figure 7-22e–j). Before dissipating late in 2011, it caused the temperature in Saturn’s atmosphere over it to soar 80 K (120°F) above normal and also led to the emission of a vast quantity of ethylene, a colorless, odorless gas. The source of that gas in Saturn is at present unknown. In 2004, *Cassini* detected lightning in Saturn’s atmosphere.

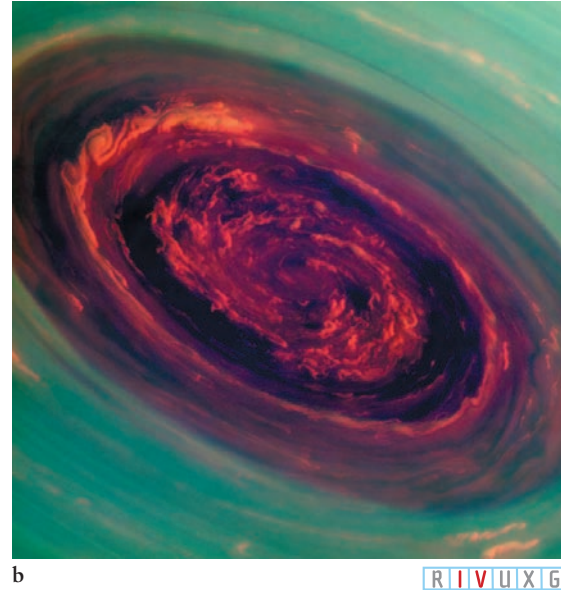
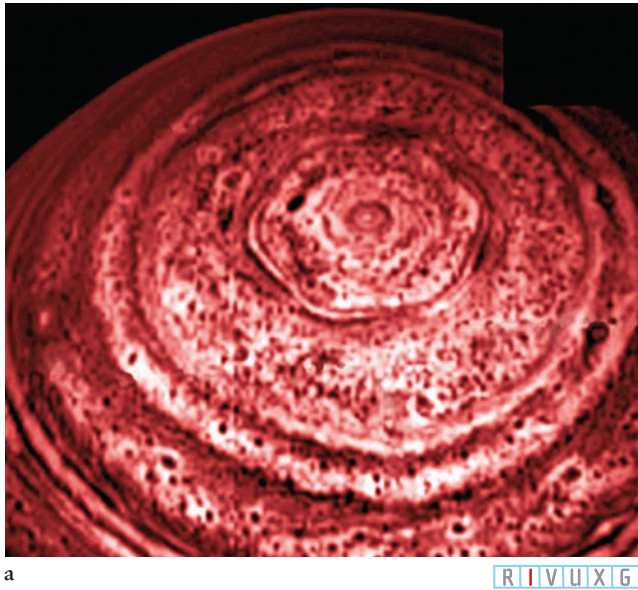


FIGURE 7-21 Belts and Zones on Saturn (a) Observing in the infrared, *Cassini* took this view of a hexagonal pattern of clouds that rotates much more slowly than the surrounding belts and zones. The pattern's origin is still under

investigation. (b) Observed in visible light, the north pole is covered by and encircled by huge storms, imaged by *Cassini* in 2012. (a: NASA/JPL/Space Science Institute; b: NASA/JPL-Caltech/SSI)

Saturn's atmosphere is composed of the same gases as Jupiter's atmosphere. However, because its mass is smaller than Jupiter's, Saturn's gravitational force on its atmosphere is less. Therefore, Saturn's atmosphere is more spread out than that of its larger neighbor (see Figure 7-6). Saturn has impressive surface winds. At its equator, storms have been clocked moving at speeds of

1600 km/h (1000 mi/h), which is 10 times faster than Earth's jet streams and 3 times faster than the fastest winds on Jupiter. The reason for Saturn's extremely high wind speeds is still under investigation.

As noted previously, the chemical composition of Saturn's atmosphere is different from that of Jupiter: Saturn's atmosphere is about 96.3% hydrogen and 3.3%

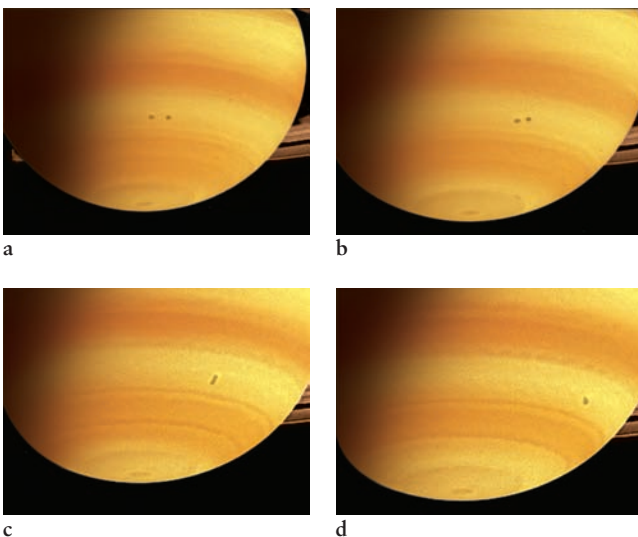
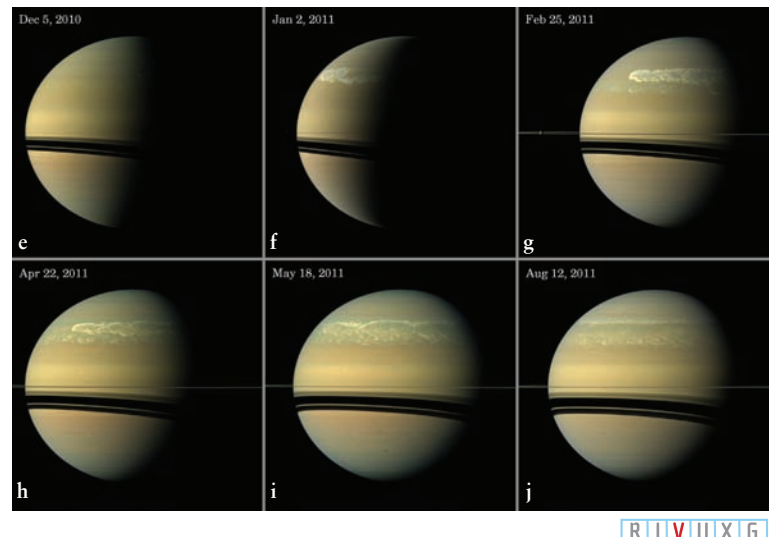


FIGURE 7-22 Merging Storms on Saturn (a–d) This sequence of *Cassini* images shows two hurricane-like storms merging into one on Saturn in 2004. Each storm is about 1000 km (600 mi)



across. (e–j) This storm swept around Saturn's northern mid-latitudes throughout 2011. (a–d: NASA/JPL/Space Science Institute. e–j: NASA/JPL-Caltech/Space Science Institute)

helium, whereas Jupiter's atmosphere is about 89.8% hydrogen and 10.2% helium. A related observation is that Saturn emits about 2.3 times as much energy as it receives from the Sun. Unlike Jupiter, whose extra energy is emitted primarily by the Kelvin-Helmholtz mechanism, much of Saturn's extra energy is believed to be generated by a mechanism called *helium rain*. Helium rain works like this: Because it is smaller than Jupiter and farther from the Sun, Saturn's atmosphere is cool enough for the helium in it to condense into liquid droplets that then descend deep into the planet. As the helium rain strikes the interior, the rain's energy helps heat the planet; the planet then in turn radiates that energy.

Astronomers infer that Saturn's interior structure resembles that of Jupiter. A layer of molecular hydrogen just below the clouds surrounds a layer of liquid metallic hydrogen and a rock and metal core (see Figure 7-8).

Because of its smaller mass, Saturn's interior is also less compressed than Jupiter's, so the pressure inside Saturn is insufficient to convert as much hydrogen into liquid metal. Its lower temperature and pressure, compared to Jupiter, prevent Saturn's core from dissolving.

The radius of Saturn's rocky core is about 7400 km (4600 mi) while its layer of liquid metallic hydrogen is 12,000 km (7460 mi) thick (see Figure 7-8). To alchemists, Saturn was once associated with the extremely dense element lead. This is wonderfully ironic in that at 687 kg/m^3 (1160 lb/yd^3), Saturn is, on average, the least dense body in the entire solar system; an object with that density on Earth would float on water.

7-10 Saturn's spectacular rings are composed of fragments of ice and ice-coated rock

Saturn's magnificent rings are among the most interesting objects in the solar system. They are tilted 27° from the perpendicular to Saturn's plane of orbit, so sometimes they are nearly edge-on to us, making them virtually impossible to see from Earth (Figure 7-23). In 1675, Giovanni Cassini discovered an intriguing feature—a dark division in the rings. Called the **Cassini division**, this gap of 5000 km (3100 mi) separates the dimmer **A ring** from the brighter **B ring**, which lies closer to the

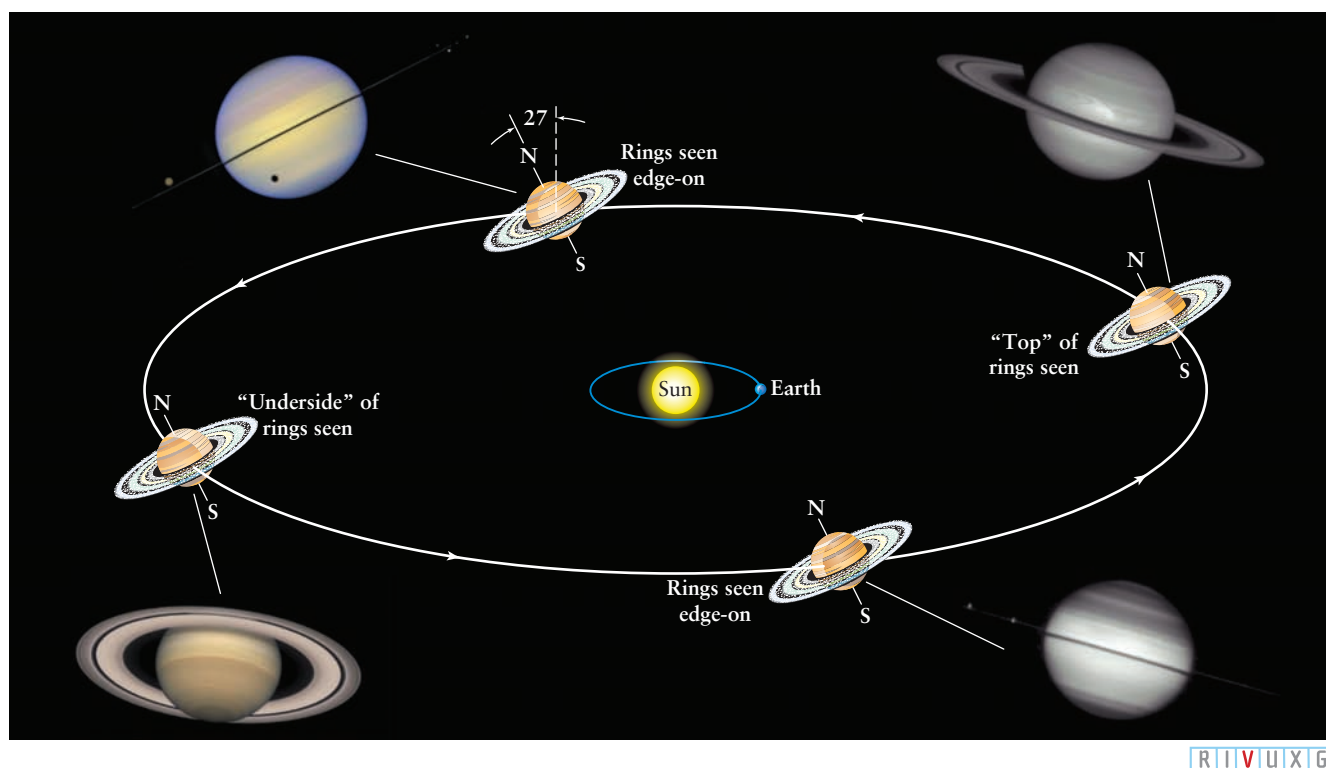


FIGURE 7-23 Saturn as Seen from Earth Saturn's rings are aligned with its equator, which is tilted 27° from the plane of Saturn's orbit around the Sun. Therefore, Earth-based observers see the rings at various angles as Saturn orbits the Sun. The plane of Saturn's rings and equator keeps the same orientation in space as the planet goes around its orbit, just as

Earth keeps its $23\frac{1}{2}^\circ$ tilt as it orbits the Sun. The accompanying Earth-based photographs show how the rings seem to disappear entirely about every 15 years (top left: E. Karkoschka/University of Arizona Lunar and Planetary Lab and NASA; bottom left: AURA/STScI/NASA; bottom right and top right: A. Bosh/Lowell Observatory and NASA)

planet. By the mid-1800s, astronomers using improved telescopes detected a faint **C ring** just inside the B ring. **Figure 7-24** shows a modern perspective of these rings.

The Cassini division occurs because the gravitational force from Saturn's moon Mimas combines with the gravitational force from the planet to keep the region relatively clear of debris. Whenever matter drifts into the

Cassini division, Mimas (orbiting at a different rate than the matter in the Cassini division) periodically exerts a force on this matter, eventually pulling most of it out of the division. This effect is called a **resonance** and is similar to what happens when you repeatedly push someone on a swing at the right time, giving them energy that enables them to go higher and higher.

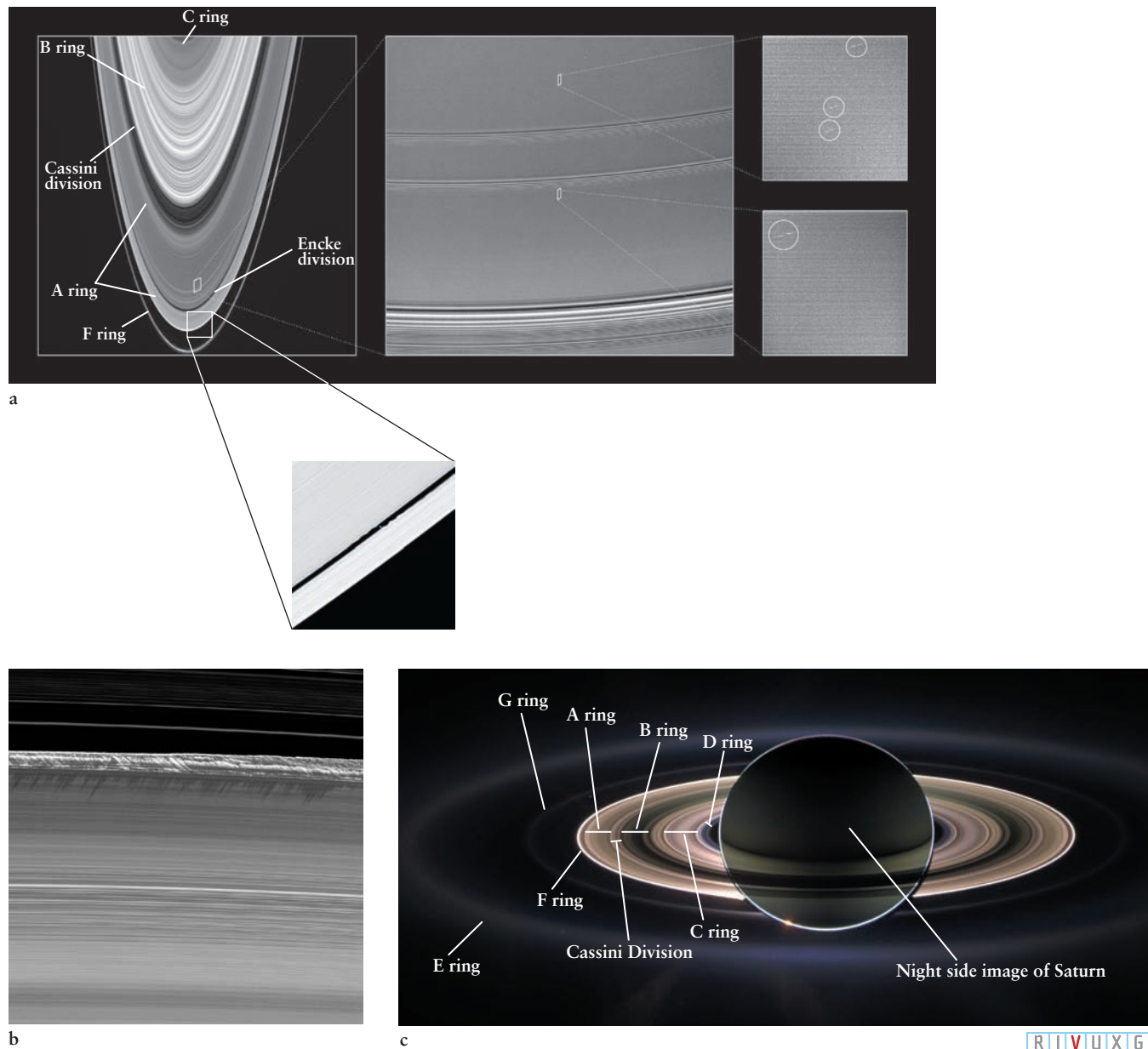


FIGURE 7-24 Numerous Thin Ringlets Constitute Saturn's Inner Rings (a) This *Cassini* image shows some of the structure of Saturn's rings, including some of the moonlets orbiting in them. **Lower inset:** As moons orbit near or between rings, they often cause the ring ices to develop ripples, like the grooves in a phonograph record. **Three insets on right:** Circles indicate four moonlets in the A ring. (b) Debris piles up along the outer edge

of Saturn's B ring just inside the Cassini division. Note the shadows of this material, indicating how irregular this edge is compared to the body of the ring. (c) Viewing Saturn's night side, the smaller pieces of debris scatter light from the Sun forward toward the *Cassini* spacecraft, including dust-sized particles in the Cassini division. Note that the Cassini division is bright in this image, while the B ring is quite dark. (NASA/JPL/Space Science Institute)

Despite the resonance keeping the Cassini division relatively clear of larger (boulder-sized) debris, it is filled with dust-sized particles. We know this because, like particles near a window in a room with sunlight streaming in, such dust-sized particles in the Cassini division scatter light *forward* rather than backward as larger pieces of matter do. We see everything around us because these things scatter light back off their surfaces toward us. Dust primarily scatters forward. When *Cassini* passed Saturn and looked back at its dark side, the dust particles in the Cassini division and the Encke division, discussed below, scattered light forward, toward the spacecraft, so that the divisions were not dark as seen from that perspective (Figure 7-24c). Further evidence for the existence of tiny debris particles in Saturn's rings where we see nothing from Earth came when spacecraft *Cassini* passed through a gap in the ring plane just outside the F ring (see Figure 7-24c). During that passage, it was peppered with 100,000 impacts from these dust particles in less than 5 minutes.

The outer edge of the B ring, just inside the Cassini division, has shown interesting structure in an image taken by the *Cassini* spacecraft. While most of the ring is only 10 m (30 ft) thick, debris piling up along that edge pushes above and below the plane of the disk (Figure 7-24b), creating a hem running around the ring, like the hem around the bottom of a skirt. This thickened part of the ring has peaks of debris ranging up to 2.5 km (1.6 mi) above the plane of the B ring. It is apparently caused by ring material flowing toward the edge of the ring and then, unable to enter the Cassini division, it flows perpendicular to the ring.

A second gap exists in the outer portion of the A ring, named the **Encke division** (Figure 7-24 and Figure 7-25), after the German astronomer Johann Franz Encke, who allegedly saw it in 1838. (Many astronomers have argued that Encke's report was erroneous because his telescope was inadequate to resolve such a narrow gap.) The first undisputed observation of the division (measuring 270 km or 170 mi in width) was made by the American astronomer James Keeler in the late 1880s, with the newly constructed 36-in (0.91-m) refractor at the Lick Observatory in California. A 40-km (25-mi) space near the outer edge of the bright A ring is now named the *Keeler gap* in honor of his work (Figure 7-26).

Unlike the Cassini division, the Encke division is kept clear because a small moon, Pan, orbits within it (Figure 7-25). The gravitational tugs of Pan, Daphnis, Prometheus, and other moons also cause the rings to ripple, as shown in the lower inset of Figure 7-24a.

The *Voyager* cameras sent back the first high-quality pictures of the **F ring**, a thin set of ringlets just beyond the outer edge of the A ring (Figure 7-26). Two tiny moons, Prometheus and Pandora, have orbits on

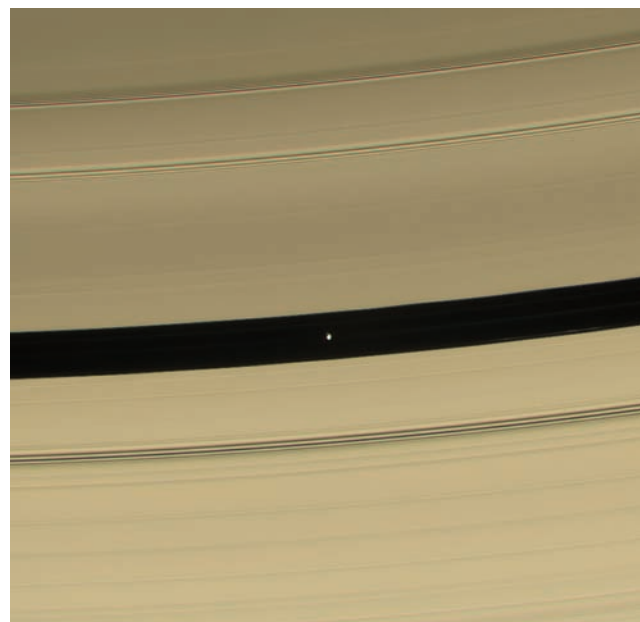


FIGURE 7-25 The Moon Pan Orbiting in the Encke division, Pan is the “shepherd” moon that keeps the division clear of small debris. It is the innermost known moon of Saturn. (NASA)

either side of the F ring and serve to keep it intact. The outer of the two satellites orbits Saturn at a slower speed than do the ice particles in the ring, as governed by Kepler's third law. As the ring particles pass near it, they receive a tiny, backward gravitational tug, which slows them down, causing them to fall into orbits a bit closer to Saturn. Meanwhile, the inner satellite orbits the planet faster than the F ring particles. This moon's gravitational force pulls them forward and nudges the particles back into a higher orbit. The combined effect of these two satellites focuses the icy particles into a well-defined, narrow band about 100 km (62 mi) wide. In 2010, the *Cassini* spacecraft observed snowballs as large as 20 km (12 mi) in diameter forming in the F ring. These bodies form as small bits of the F ring debris swirl around and collide with each other under the gravitational attractive force of the passing moon Prometheus. At other times, large pieces collide and break into smaller ones.

Because of their confining influence, Prometheus and Pandora are called **shepherd satellites** or **shepherd moons**. Among the most curious features of the F ring is that the ringlets are sometimes braided or intertwined and sometimes separate. Although stabilized by the shepherd moons, the F ring does ripple and shows varying brightness. Astronomers theorize that the changes in brightness result from collisions between clumps of matter in the ring.

Focus Question 7-9

Why are there gaps in Saturn's rings?

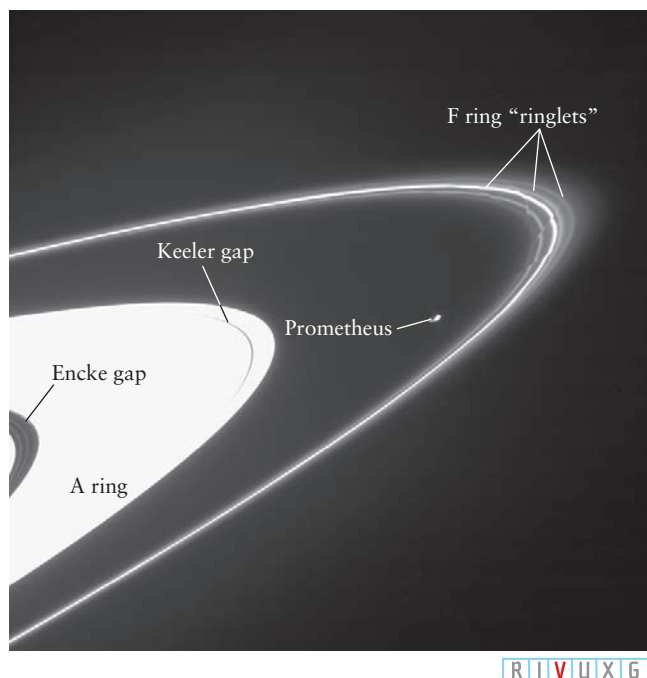


FIGURE 7-26 The F Ring and One of Its Shepherds Two tiny satellites, Prometheus and Pandora, each measuring about 50 km (31 mi) across, orbit Saturn on either side of the F ring. Sometimes the ringlets are braided, sometimes parallel to each other. In any case, the passage of the shepherd moons causes ripples in the rings, which lead to the formation of large snowballs in them. The gravitational effects of these two shepherd satellites confine the particles in the F ring to a band about 100 km (62 mi) wide. (NASA)

A promising theory has been put forth to explain the origin of the ice-rich rings of Saturn that we have just been discussing. Just as the planets formed in a disk around the protosun, astronomers believe that Saturn's largest moons, including Rhea, Titan, and Iapetus, formed in a disk of gas and dust orbiting the young Saturn. (We discuss these moons shortly.) Another moon is posited to have formed at the same time in an orbit closer to Saturn than these moons. Recall from the Nice model (Section 4-3) that Saturn formed during a period in which there were myriad pieces of debris in its orbit and crossing its orbit. That debris took energy away from Saturn, which is why it initially spiraled sunward. The theory proposes that the innermost moon, buffeted by the same debris, lost energy and spiraled inward toward Saturn. Eventually, it crossed the Roche limit, the distance at which a planet creates tides on a moon's solid surface high enough to pull the moon apart. The tidal force stressed the moon, causing much of the ice-rich surface to be stripped off and inserted into orbit as the denser rocky debris continued spiraling inward until it sank into the planet. The remaining icy debris became the major rings of Saturn. As one would expect from this theory, these rings are all inside Saturn's Roche limit.

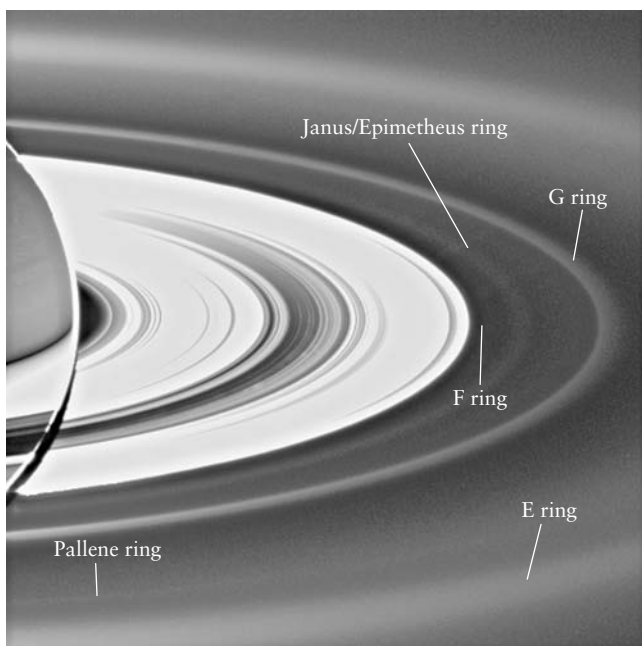
Saturn's rings are being struck by small pieces of space debris, or meteors, ranging in size from a centimeter to several meters across. These events, which occur several times every day, create temporary streaks or ripples across the rings. It appears that many of these meteors are broken up by impacts with ring debris, after which they become new ring matter.

Beyond the major rings described here and well outside Saturn's Roche limit lie extensive rings of very fine dust particles, typically smaller than a millionth of a meter (a micron) across. Because such rings are very unstable (that is, the particles are drifting away from each other), there must be small moons that stabilize the rings and replace lost debris. The moons Janus, Epimetheus, and Pallene are associated with these rings. Their gravitational attractions slow the loss of the dust while also providing replacement particles.

Figure 7-27a shows a *Cassini* image of part of this system, including a ring discovered in the orbit of the moon Pallene. The *Cassini* image in Figure 7-27b gives us a more comprehensive view of Saturn's ring system. Nevertheless, the rings go out farther still. The ring system shown in Figure 7-27c actually extends out to Titan, 1.2 million km (750,000 mi) from Saturn. In 2009, astronomers using the Spitzer Space Telescope observed a separate ring in the same orbit as the moon Phoebe (Figure 7-28), 13 million km (8.1 million mi) from Saturn. Called the *Phoebe ring*, its particles are believed to come from that moon, probably ejected as a result of meteorite impacts. Unlike the main ring system, which orbits over Saturn's equator, the Phoebe ring is inclined 27° from the equator.

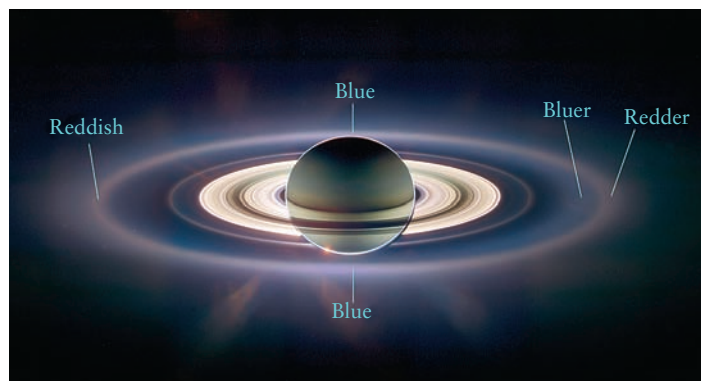
Returning to the inner rings, pictures from the *Cassini* spacecraft and the Hubble Space Telescope show that the bulk of Saturn's rings lies in a remarkably thin disk—about 10 m (33 ft) thick according to recent observations! This value is amazing when you consider that the total ring system has a width (from inner edge to outer edge) of over 380,000 km (236,000 mi), more than the distance from Earth to our Moon. In 2009, when the rings were seen edge-on from Earth, astronomers discovered that small amounts of ring material extend up to 18,000 km (11,200 mi) above and below the plane of the rings, probably pulled out by the gravitational attraction of Saturn's moons and by the planet's magnetic field. In 2006, *Cassini* discovered a new class of small moons, called *moonlets*, orbiting in Saturn's rings. As many as 10 million of them may exist there. One theory is that they are pieces of a larger body (see Figure 7-24) that broke up to form the rings.

The particles that form Saturn's rings must be highly reflective because the rings are very bright (their albedo is 0.80). Astronomers had long suspected that the rings consist primarily of water ice, along with ice-coated rocks. Spectra taken by Earth-based observatories and by the spacecraft visiting Saturn have confirmed this



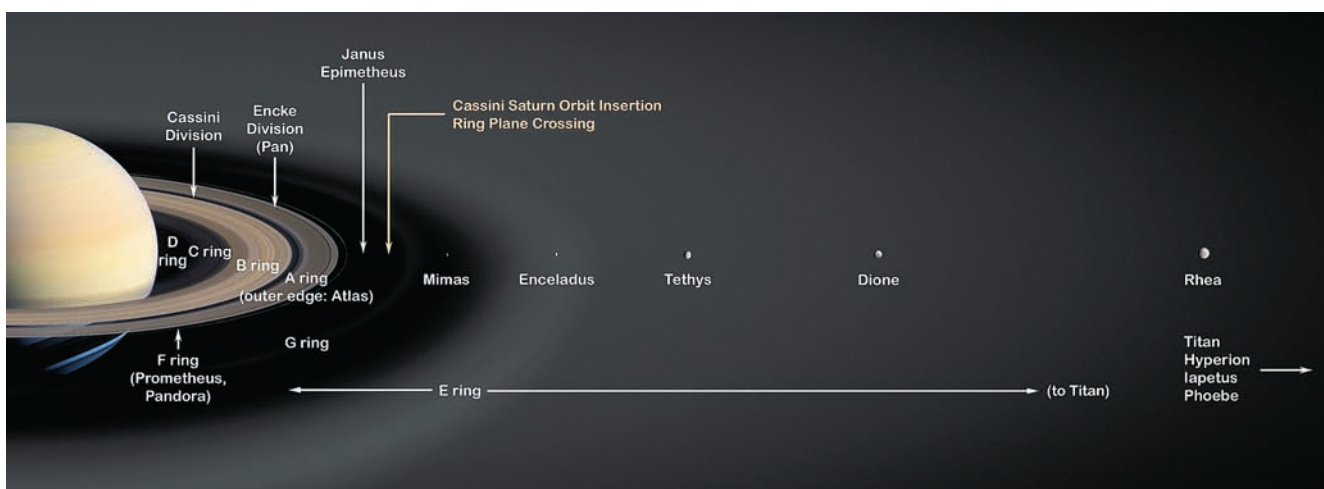
a

R I V U X G



b

R I V U X G



c

R I V U X G

FIGURE 7-27 Saturn's Ring System (a) Photographed with the Sun behind Saturn, the inner and intermediate regions of Saturn's ring system are shown to be very different from each other. Beyond the F ring, the particles are dust and pebble-sized. (b) Another view of the inner and intermediate rings,

where subtle color differences are indicated. (c) Superimposed on this *Cassini* image are labels that indicate how far the rings extend into the moon system of Saturn. Titan (off image on right) is 1.2 million km (750,000 mi) from the center of Saturn. (a, b: NASA/JPL/Space Science Institute; c: NASA/JPL)

suspicion. While the temperature of the rings ranges from 93 K (−90°F) in the sunshine to less than 73 K (−330°F) in Saturn's shadow, water particles are nevertheless being freed from them, probably by ultraviolet radiation from the Sun. The water particles are guided toward Saturn by the planet's magnetic fields, where they combine with other particles in Saturn's upper atmosphere.

Saturn's rings are slightly salmon-colored. This coloring suggests that they contain traces of organic molecules, which often have such hues. It appears that the

rings have gained this material by being bombarded with debris from the outer solar system.

To determine the size of the particles in Saturn's rings, scientists measured the brightness of different rings from many angles as the *Voyager* and *Cassini* spacecraft flew past the planet. Astronomers also measured changes in radio signals received from these spacecraft as they passed behind the rings. Different rings are composed of debris with different sizes, ranging from particles of the sizes found in wood smoke, through those a few centimeters across, to particles roughly 10 m (11 yd) in

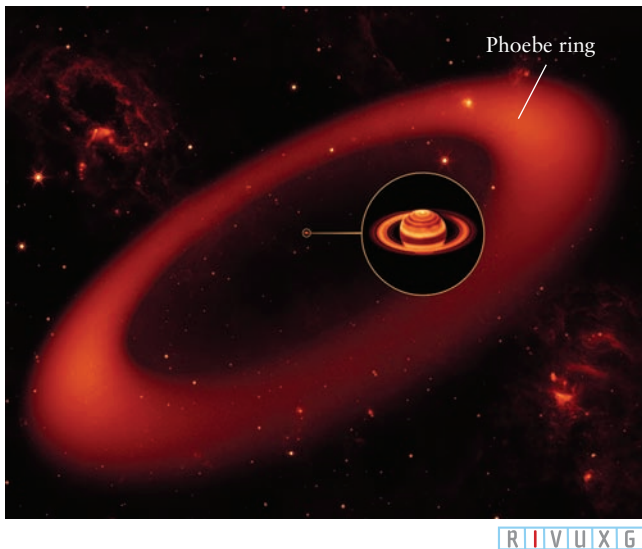


FIGURE 7-28 Saturn and Its Outer Giant Ring This artist's rendition of Saturn's giant ring is drawn to scale with an infrared image of Saturn and the rings we normally see. The giant ring spans the region from 6 million km (3.7 million mi) to 18 million km (11.1 million mi) beyond Saturn. Put another way, the giant ring is as wide as 30 Saturns placed side by side. (NASA/JPL-Caltech/Keck)

diameter. The outermost main ring, the A ring (see Figure 7-24), is composed of these larger clumps of rubble, which frequently collide with one another and thereby change sizes. This ongoing process is analogous to the planet-formation activity described in Chapter 4. Of the particles that would be visible to the naked eye (that is, excluding dust particles), the centimeter-sized particles are most common.

5 High-resolution images reveal that the ring structure as seen from Earth actually consists of thousands of closely spaced ringlets (**Figure 7-29**).

Saturn's layer of liquid metallic hydrogen produces a planetwide magnetic field that affects its rings. Due to Saturn's much smaller volume of liquid metallic hydrogen, the magnetic field at Saturn's surface is only about two-thirds as strong as the magnetic field here on Earth. Data from spacecraft show that Saturn's magnetosphere contains radiation belts similar to Earth's Van Allen belts. Furthermore, dark spokes move around Saturn's rings (see Figure 7-29); these are believed to be created by electric charges on the ring material interacting with the planet's magnetic field. The magnetic field lifts charged particles out of the plane in which the rings orbit. Spreading the particles out decreases the light scattered from them and therefore makes the rings appear darker. The magnetic field, moving with the planet as it rotates, causes the regions of spread particles to change, making the spokes appear to revolve around the planet.

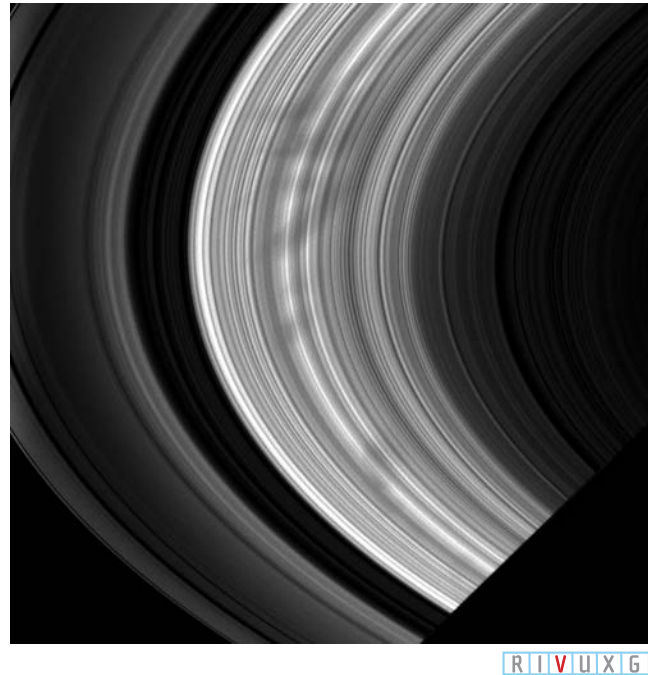


FIGURE 7-29 Spokes in Saturn's Rings Believed to be caused by Saturn's magnetic field moving electrically charged particles that are lifted out of the ring plane, these dark regions move around the rings like the spokes on a rotating wheel. (NASA/JPL/Space Science Institute)

7-11 Titan has a thick atmosphere, clouds, and lakes filled with liquids

Only seven of Saturn's 62 known moons are spherical. Four of them are shown in **Figure 7-30a–c, e**. The other 55 moons are oblong, suggesting that they are captured asteroids. These oblong satellites include Phoebe (Figure 7-30d) and Hyperion (Figure 7-30f). Twelve oblong moons discovered in 2001 move in clumps, suggesting that they are pieces of a larger body that was broken up by impacts.

Saturn's largest moon, Titan, is second in size to Ganymede among the moons of the solar system and is the only moon in the solar system found to have a dense atmosphere. Like many of the other moons of Saturn, Titan is in synchronous rotation around the planet. About 10 times more gas lies above each square meter of Titan's surface than lies above each square meter of Earth. As a result, the air pressure on the surface of Titan is about 1.5 times the air pressure we feel on Earth.

Christiaan Huygens discovered Titan in 1655, the same year he proposed that Saturn has rings. By the early 1900s, several scientists had begun to suspect that Titan might have an atmosphere. This moon is larger (although less massive) than Mercury and far colder than that planet. Calculations reveal that Titan is cool enough and massive enough to retain heavy gases in its atmosphere. Because of this, Titan was a primary target

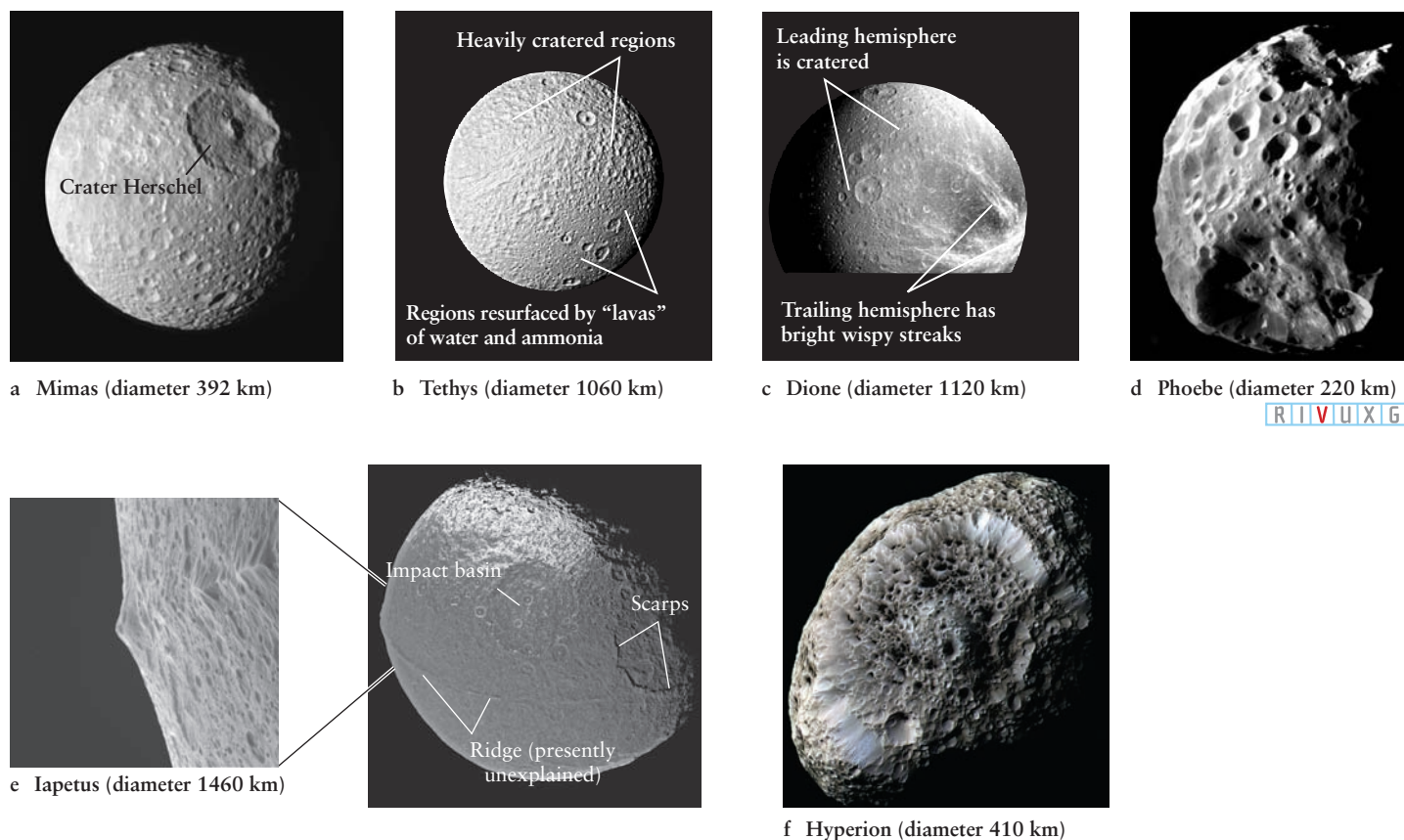


FIGURE 7-30 Saturn's Diverse Moons (a–c) These *Voyager 1*, *Voyager 2*, and *Cassini* images show the variety of surface features seen on three of Saturn's seven spherical moons. They are not shown to scale (refer to the diameters given below each image). (d) In comparison, this *Cassini* image shows the nonspherical moon Phoebe, almost as dark as coal, carrying many craters, landslides, grooves, and ridges. Phoebe is barely held in orbit by Saturn. Astronomers believe that it was captured after wandering in from beyond the orbit of Neptune. Two particularly intriguing moons are (e) Iapetus and (f)

Hyperion. The ridge running along the equator of Iapetus is believed to have developed as the moon formed. Apparently Iapetus cooled so rapidly that the ridge did not have time to settle away. Perhaps the most bizarre object photographed in the solar system, ice-covered Hyperion, shows innumerable impact craters. These features are different from craters seen in other objects in that the crater walls here have not filled in the bottom of the craters. This moon's low gravity and the pull of nearby Titan may explain this unusual phenomenon. (a, d–f: NASA/JPL/Space Science Institute; b, c: JPL/NASA)

for the *Voyager* missions (Figure 7-31a). Unexpectedly, its thick haze completely blocked any view of its surface by *Voyager*'s relatively primitive cameras. Observations also reveal that clouds of methane form and dissipate seasonally near both of Titan's poles and in the mid-southern latitudes. Although the causes for the appearance and disappearance of these clouds are not all known, geysers and evaporation from Titan's lakes are considered as possible factors.

Voyager data indicated that roughly 90% of Titan's atmosphere is nitrogen. Most of this gas probably formed from the breakdown of ammonia (NH_3) into hydrogen and nitrogen atoms under the influence of the Sun's ultraviolet radiation. Because Titan's gravity is too weak to retain hydrogen, this gas has escaped into space, leaving behind ample nitrogen. The unbreathable atmosphere is about 4 times as dense as that of Earth.

The second most abundant gas on Titan is methane, a major component of natural gas. Sunlight interacting

with methane induces chemical reactions that produce a variety of other carbon-hydrogen compounds, or **hydrocarbons**. Spacecraft have detected small amounts of nearly 20 hydrocarbons in Titan's atmosphere, including ethane (C_2H_6), acetylene (C_2H_2), ethylene (C_2H_4), and propane (C_3H_8).

Methane and ethane condense into droplets in the cold air, making Titan the only other world in the solar system besides Earth where rain reaches the surface. The rain, however, is very different on the two worlds. Titan's raindrops are calculated to be 1 cm (0.4 in) in diameter (roughly twice the size of large raindrops on Earth) and, because of Titan's low surface gravity, they fall 6 times more slowly than raindrops do here—rain on Titan falls at the speed of slowly falling snowflakes on Earth. Nitrogen combines with various hydrocarbons to produce other compounds that can exist as liquids on Titan. Although one of these compounds, hydrogen cyanide (HCN), is a poison, some of the others are the building

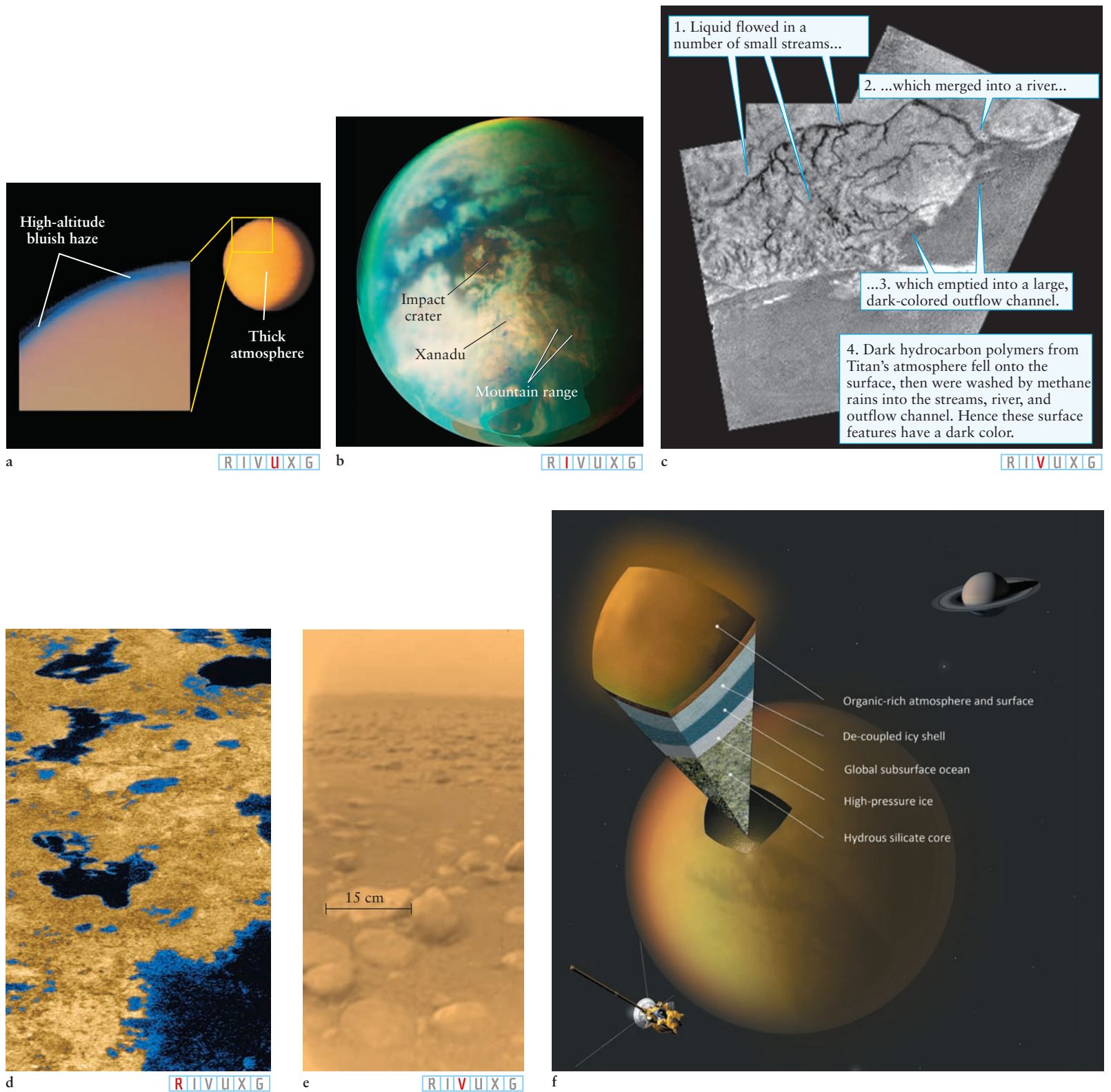


FIGURE 7-31 Surface Features on Titan (a) *Voyager* images of Titan's smoggy atmosphere. (b) *Cassini* image of Titan (diameter 5150 km or 3200 mi) showing lighter highlands, called Xanadu, and dark, flat, lowlands that may be hydrocarbon seas. Resolution is 4.2 km (2.6 mi). (c) Riverbeds meandering across the Xanadu highlands of Titan. These are believed to have been formed by the flow of liquid methane and ethane. (d) Lakes filled with liquid methane and ethane. (e) The *Huygens* probe

took this image at Titan's surface on January 14, 2005. What appear like boulders here are actually pebbles strewn around the landscape. The biggest ones are about 15 cm (6 in.) across. (f) Our present understanding of Titan's structure. (a: NASA/JPL/Space Science Institute; b: NASA/JPL/University of Arizona; c, e: NASA/JPL/ESA/University of Arizona; d: Cassini Radar Mapper, JPL, ESA, NASA; f: NASA/JPL-Caltech)

blocks of life's organic molecules. Indeed, biologists who study Titan's atmosphere now think that its chemistry is very similar to that of the young Earth.

Focus Question 7-10

What liquids appear to exist on Titan?

Many of the hydrocarbons and carbon-nitrogen compounds in Titan's atmosphere can form complex compounds called **polymers**. These molecules are long repeating chains of atoms; rubber, cellulose, and plastics are the best-known examples on Earth. Scientists hypothesize that droplets of lighter polymers remain suspended in Titan's atmosphere and form a mist, while heavier polymer particles settle down onto Titan's surface.

Images of the surface taken by the *Cassini* spacecraft (Figure 7-31b), which arrived in the Saturn system in 2004, show few craters, indicating that Titan's surface undergoes dynamic change. Despite the fact that activity in its atmosphere and surface continually covers craters and other surface features, Titan does have a (presumably long-lived) mountain range with peaks as high as 1.5 km (1 mi). These peaks are covered with organic matter and possible methane snow.

About 20% of Titan's surface has sand dunes 100 m (109 yd) high. The composition of the sand is not yet clear, although it may be made of frozen organic molecules. Using the Keck I telescope in 1999, astronomers observed infrared emissions from Titan's surface consistent with a partially liquid surface. Rain and rivers of methane have been observed (Figure 7-31c). In 2007, radar images from *Cassini* revealed a smooth dark surface of over 10^5 km² (39,000 mi²), the largest lake of liquid methane and ethane seen on Titan. Other images, such as Figure 7-31d, reveal similar features elsewhere on that world, including islands in some lakes. Titan's lakes are all located near its poles.

Cassini launched a probe named *Huygens* that successfully parachuted onto Titan and photographed rocks

(Figure 7-31e), highlands, and channels. Upon impact, *Huygens* vaporized methane that had been just centimeters below the surface. The quantity of methane and ethane detected there implies that Titan has reservoirs of these and other hydrocarbons, which are continually being broken down in its atmosphere by the Sun's ultraviolet radiation.

We have little reason to suspect that life exists on Titan's surface because the temperature there, 95 K (−288°F), is prohibitively cold. However, in 2008 astronomers using data from *Cassini* discovered that surface features, including lakes, canyons, even mountains, drift some 15 km (9.3 mi) per year. They concluded that this drift is possible because a layer of liquid water and ammonia exists 100 km (62 mi) under Titan's surface. That layer acts as a lubricant that allows the surface to move relative to the deep interior, which has been determined to be a mix of rock and ice. The combination of liquid water and organic compounds, such as methane, may enable life to have evolved inside Titan. Our present understanding of Titan's interior structure is summarized in Figure 7-31f.

7-12 Rhea has ice

Rhea, Saturn's second-largest moon (Figure 7-32a), is also in synchronous rotation. Astronomers have observed that its leading face has many more craters than its trailing face. This moon has a density consistent with a composition of mostly water ice mixed with rocky debris. In 2008, *Cassini* observed white and blue ice features on its trailing side that are associated with scarps (long lines of cliffs) like those found on Mercury. In 2010, *Cassini* detected an atmosphere containing 70% oxygen and 30% carbon dioxide, which is about 100 times thicker than the atmosphere of Mercury.

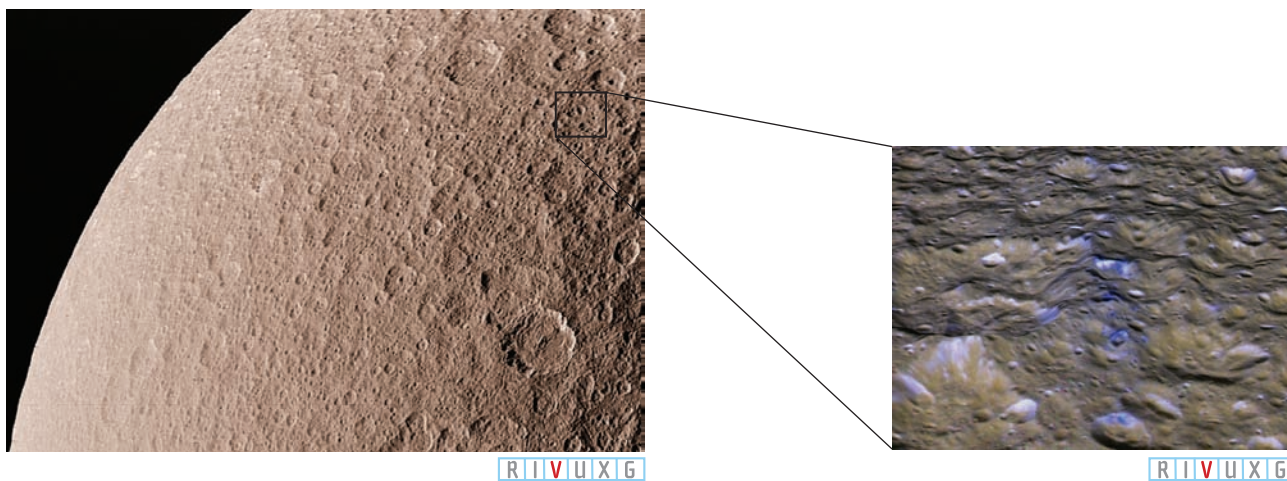


FIGURE 7-32 Rhea Rhea (diameter 1530 km or 950 mi) is heavily cratered. The bluish regions on the inset are believed to be ices uncovered as a result of impacts. (Voyager, NASA;

Copyright: © 2002 Calvin J. Hamilton; inset: NASA/JPL/Space Science Institute/Universities Space Research Association/Lunar & Planetary Institute)

7-13 Enceladus has water jets, an atmosphere, and a magnetic field

Saturn's sixth-largest moon, Enceladus (500 km [310 mi] in diameter, **Figure 7-33a**), moved from the sidelines toward center stage in 2005 as *Cassini* began studying it. This bright world has an icy, wrinkled surface similar in appearance to Jupiter's Europa and Ganymede. The fact that these latter two bodies contain liquid water was the first clue that Enceladus does, too. This conclusion is supported by observations that entire regions of Enceladus's southern hemisphere are free of craters. Instead, there are ice blocks the size of large houses, which have apparently erased many craters due to tectonic plate motion (Figure 7-33b). In addition, the spacecraft revealed that dark lines on Enceladus, dubbed *tiger stripes* (Figure 7-33a), are less than 1000 years old.

These features, much warmer than the surrounding surface and heated from below, are cracks through which water wells up from inside, freezing into a smooth, dark surface. The tiger stripes are typically 100 km (62 mi) long and 35 km (22 mi) apart.

Proof positive that Enceladus has liquid water inside it came in 2007 when jets of water were first seen shooting out of it at speeds up to 2200 km/h (1400 mi/h) from the regions of the tiger stripes (Figure 7-33c). The most promising explanation for the origin of the liquid water and the cause of the recent formation of the tiger stripes is the heating of Enceladus by the tidal force from Saturn. Similar to how Jupiter's and Europa's gravitational forces cause Io to change shape and thereby heat

Focus Question 7-11

Besides Mercury, what other world that we have studied has scarps?

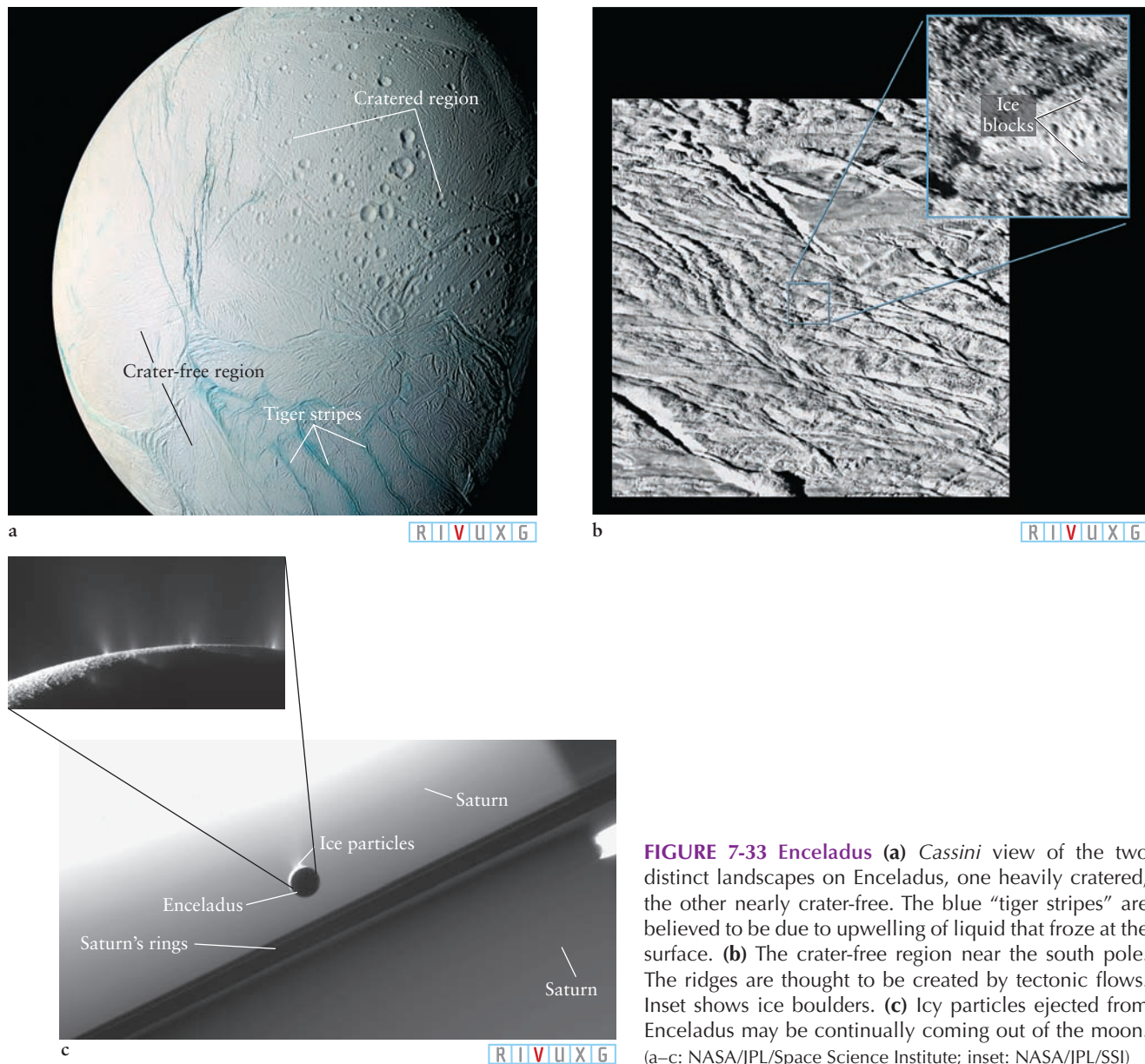


FIGURE 7-33 Enceladus (a) *Cassini* view of the two distinct landscapes on Enceladus, one heavily cratered, the other nearly crater-free. The blue “tiger stripes” are believed to be due to upwelling of liquid that froze at the surface. (b) The crater-free region near the south pole. The ridges are thought to be created by tectonic flows. Inset shows ice boulders. (c) Icy particles ejected from Enceladus may be continually coming out of the moon. (a–c: NASA/JPL/Space Science Institute; inset: NASA/JPL/SSI)

as it orbits, the gravitational forces from Saturn and its host of other moons cause Enceladus to change shape. This motion creates internal friction and heat that keep a large volume of water inside Enceladus liquid, while also stressing the surface, allowing the tiger stripes to form.

Observations indicate that this water leaves Enceladus cyclically; when the moon is farthest from the planet, water emerges in relatively steady jets (unlike geysers on Earth, which send up water in spurts). When Enceladus is close to Saturn, the jets are turned off. Tides on the moon explain this behavior. When Enceladus is close to Saturn, the tidal force from the planet adjusts the moon's surface shape to stop the jets of water, whereas when Enceladus is farther from the planet, the moon's shape is altered, opening gaps through which the water is ejected. Because Enceladus lacks sufficient gravitational attraction to pull all the water it ejects down onto its icy surface, this moon is continually losing mass.

Cassini's detectors also discovered that Enceladus has a thin atmosphere and a magnetic field (consistent with salty liquid water in its interior). Besides water vapor, the atmosphere of Enceladus is composed of hydrogen, carbon dioxide, nitrogen, carbon monoxide, and organic compounds. *Cassini* also detected dust and salt-rich ice grains surrounding Enceladus. This material is the source of Saturn's E ring.

URANUS

Uranus (pronounced *YUR-uh-nus*), so far away that it is virtually invisible to the naked eye, was the first planetary system discovered by a telescopic search of the heavens. The planet is 19.2 AU from the Sun and takes 84 Earth years to orbit it. Since its discovery in 1781, Uranus has orbited the Sun fewer than 3 times.

7-14 Uranus sports a hazy atmosphere and clouds

Until early 1996, observations of Uranus, the fourth most massive planet in our solar system, revealed few notable features in the visible part of the spectrum. Even the 1986 visit of *Voyager 2* showed a remarkably bland world. It took the Hubble Space Telescope's infrared camera to find what *Voyager's* visible-light camera could not: Uranus has a system of belts and zones. Its hydrogen atmosphere has traces of methane along with a high-altitude haze, under which are clear air and ever-changing methane clouds that dwarf the typical cumulus clouds we see on Earth. Uranus's clouds

are towering and huge, each typically as large as Europe. Along with the rest of the atmosphere, the clouds go around the planet once every 16½ hours, with variation due to differential rotation, as also occurs on Jupiter and Saturn.

Uranus contains 14½ times as much mass as Earth and is 4 times bigger in diameter (Figure 7-34). Its outer layers are composed predominantly of gaseous hydrogen and helium. The temperature in the upper atmosphere of the planet is so low (about 73 K, -330°F) that the methane and water there condense to form clouds of ice crystals. Because methane freezes at a lower temperature than water, methane forms higher clouds over Uranus than does the water vapor. Methane efficiently absorbs red light while scattering blue and green, thereby giving the planet its blue-green color.

Uranus's rotation axis lies very nearly in the plane of its orbit around the Sun (Figure 7-35). Put another way, recall that Earth's rotation axis is tilted about 23½° from being perpendicular to the ecliptic. Uranus's rotation axis is inclined 98° from a line perpendicular to its plane of orbit (see lower-left drawing in Figure 7-35). Therefore, its rotation axis lies within 8° of the plane of orbit.

Because its north pole points below the plane of its orbit around the Sun, Uranus is one of only two planets with retrograde rotation (Venus is the other). As Uranus orbits the Sun, its north and south poles alternately point almost directly toward or directly away from the Sun, producing exaggerated seasons (see Figure 7-35). In the summertime, near Uranus's north pole, the Sun is almost directly overhead for many Earth years, at which time southern latitudes are subjected to a continuous, frigid winter night. Forty-two Earth years later, the situation is reversed.

From *Voyager* photographs, planetary scientists have concluded that each of the five largest Uranian moons has probably had at least one shattering impact. A catastrophic collision with an Earth-sized object may also have knocked Uranus on its side, as we see it today.

From its mass and density (1318 kg/m³ or 2222 lb/yd³), astronomers conclude that Uranus's interior has three layers. The outer 30% of the planet is liquid hydrogen and helium, the next 40% inward is highly compressed liquid water (with some methane and ammonia), and the inner 30% is a rocky core (Figure 7-36). Ammonia dissolves easily in water, so an ocean would explain the scarcity of this gas in the planet's atmosphere. Because Uranus and Neptune are so rich in water, they are called *ice giants*.

Uranus's surface magnetic field is about three-quarters that of Earth. That strength is reasonable, considering the planet's mass and rotation rate, but everything else about the magnetic field is extraordinary. The field is remarkably tilted—59° from its axis of rotation—and its axis does not even pass through the center of the planet (Figure 7-37).

The magnetosphere of Uranus wobbles considerably as the planet rotates because of the large angle between the magnetic field of Uranus and its rotation axis. A rapidly changing magnetic field like this occurs elsewhere in astronomy. For example, it helps explain the behavior of pulsars, a type of stellar remnant we

Focus Question 7-12

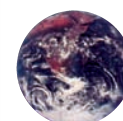
Besides the terrestrial world that is its core, what substance dominates the interior of Uranus?

URANUS: Vital Statistics

Planet symbol:	♅
Average distance from the Sun:	19.2 AU = 2.87×10^9 km = 1.78×10^9 mi
Maximum distance from the Sun:	20.0 AU = 3.00×10^9 km = 1.86×10^9 mi
Minimum distance from the Sun:	18.4 AU = 2.75×10^9 km = 1.71×10^9 mi
Orbital eccentricity:	0.047
Average orbital speed:	6.83 km/s = 4.24 mi/s
Orbital period:	84.1 years
Rotation period (internal):	17.24 h
Inclination of equator to orbit:	97.9°
Inclination of orbit to ecliptic:	0.77°
Diameter:	57,118 km = 35,500 mi = 4.01 Earth diameters (equatorial)
Mass:	8.68×10^{25} kg = 14.5 Earth masses
Average density:	1318 kg/m ³ = 2222 lb/yd ³
Escape speed:	21.3 km/s = 13.2 mi/s
Surface gravity (Earth = 1):	0.90
Albedo:	0.56
Average temperature at cloud tops:	−218°C = −360°F = 55 K
Atmospheric composition (by number of molecules):	82.5% hydrogen (H ₂), 15.2% helium (He), 2.3% methane (CH ₄)



Uranus

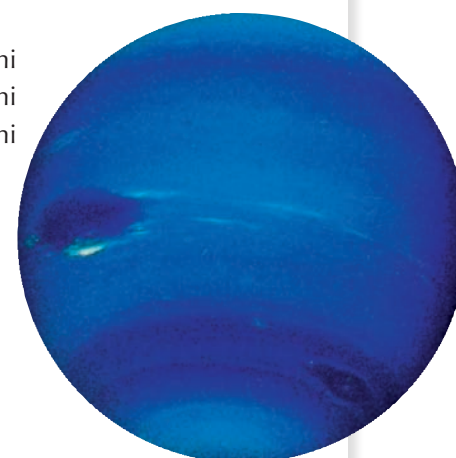


Earth

R I V U X G

NEPTUNE: Vital Statistics

Planet symbol:	♆
Average distance from the Sun:	30.1 AU = 4.50×10^9 km = 2.80×10^9 mi
Maximum distance from the Sun:	30.4 AU = 4.54×10^9 km = 2.82×10^9 mi
Minimum distance from the Sun:	29.8 AU = 4.45×10^9 km = 2.77×10^9 mi
Orbital eccentricity:	0.009
Average orbital speed:	5.5 km/s = 3.4 mi/s
Orbital period:	164.8 years
Rotation period (internal):	16.11 h
Inclination of equator to orbit:	29.6°
Inclination of orbit to ecliptic:	1.77°
Diameter:	49,528 km = 30,775 mi = 3.88 Earth diameters (equatorial)
Mass:	1.024×10^{26} kg = 17.1 Earth masses
Average density:	1638 kg/m ³ = 2761 lb/yd ³
Escape speed:	23.5 km/s = 14.6 mi/s
Surface gravity (Earth = 1):	1.1
Albedo:	0.41
Average temperature at cloud tops:	−218°C = −360°F = 55 K
Atmospheric composition (by number of molecules):	79% hydrogen (H ₂), 18% helium (He), 3% methane (CH ₄)



Neptune



Earth

R I V U X G

FIGURE 7-34 Uranus, Earth, and Neptune These images of Uranus, Earth, and Neptune are to the same scale. Uranus and Neptune are quite similar in mass, size, and chemical composition.

Both planets are surrounded by thin, dark rings, quite unlike Saturn's, which are broad and bright. The clouds on the right of Uranus (false-color pink) are each the size of Europe. (NASA)

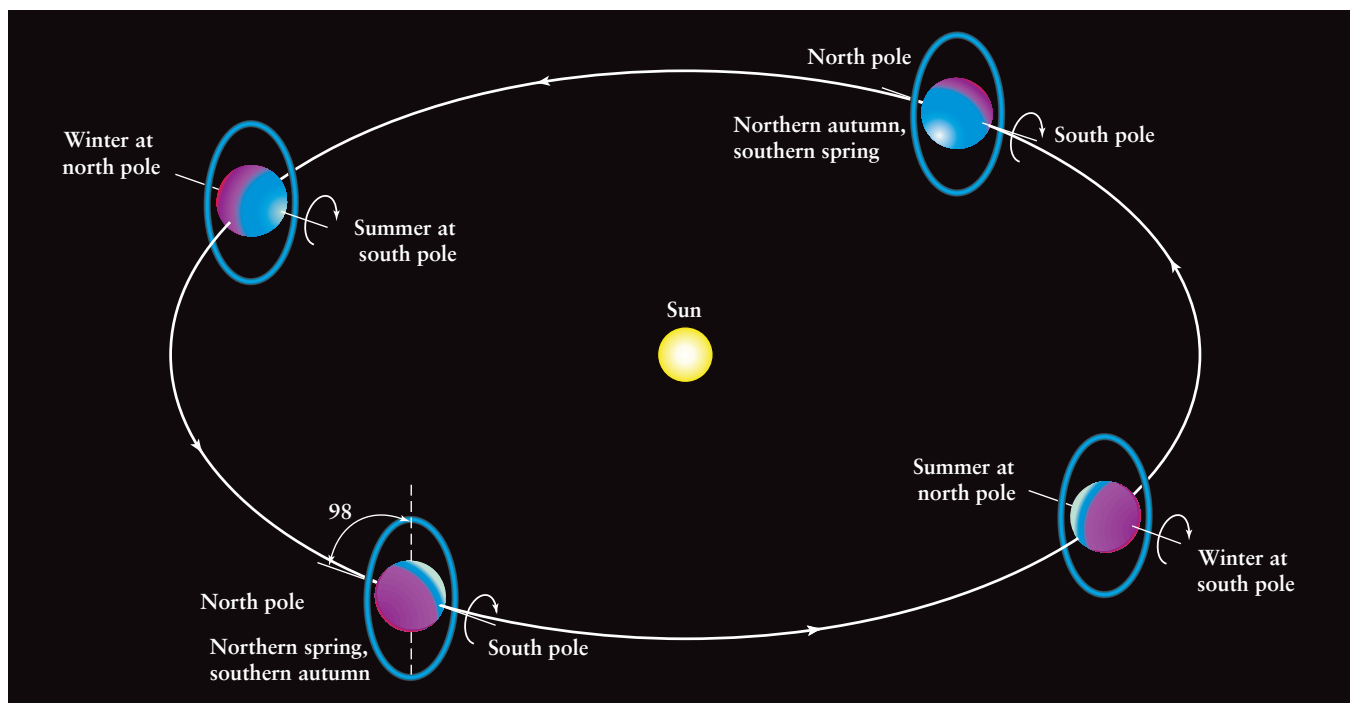


FIGURE 7-35 Exaggerated Seasons on Uranus Uranus's axis of rotation is tilted so steeply that it lies nearly in the plane of its orbit. Seasonal changes on Uranus are thus greatly exaggerated compared to those we experience. For example,

will study in Chapter 12. Astronomers use Uranus's magnetosphere's wobble to determine that the planet's interior rotates once every 17 hours 14 minutes. This is considerably slower than the 16½-hour rotation rate of the clouds on its surface.

7-15 A system of rings and satellites revolves around Uranus

Uranus has 15 known rings, each divided into ringlets (**Figure 7-38**). The first nine of the thin, dark rings were

discovered accidentally in 1977 when Uranus passed in front of a star. The star's light was momentarily blocked by each ring, thereby revealing their existence to astronomers (**Figure 7-39**). Blocking the light of a more distant object, such as the star here, by something between it and us, such as Uranus's rings, is called an **occultation**.

The other rings were detected by *Voyager* and the Hubble Space Telescope. The rings are held in orbit by shepherd moons, including Cordelia and Ophelia.

Most of Uranus's 27 known moons, like its rings, orbit in the plane of the planet's equator, that is, nearly

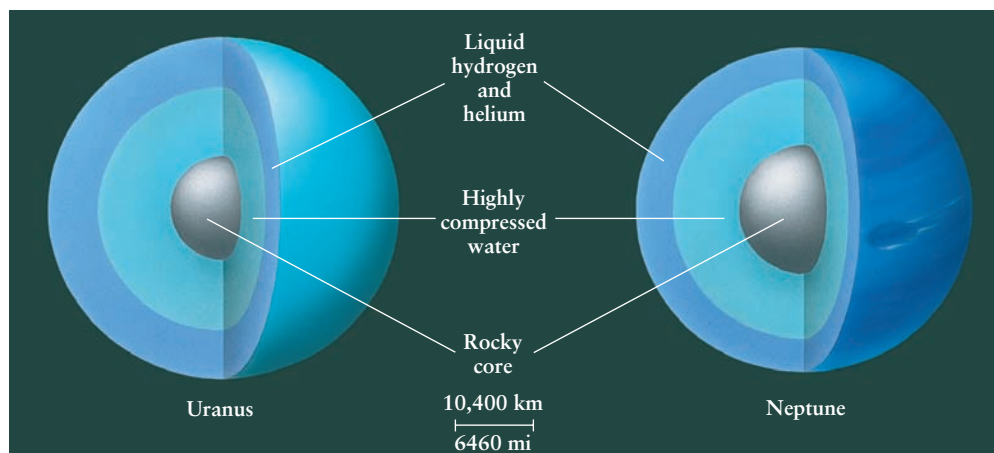


FIGURE 7-36 Cutaways of Uranus and Neptune The interiors of both Uranus and Neptune are believed to have three regions: a terrestrial rocky core surrounded by a liquid water

mantle, which is surrounded, in turn, by liquid hydrogen and helium. Their atmospheres are thin layers at the top of their hydrogen and helium layers.

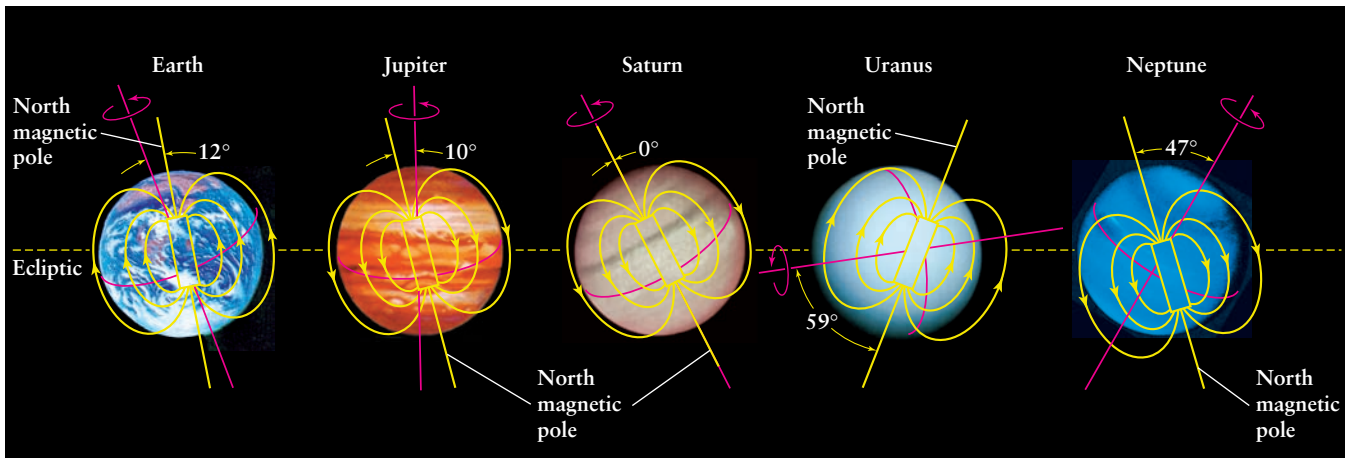


FIGURE 7-37 The Magnetic Fields of Five Planets This drawing shows how the magnetic fields of Earth, Jupiter, Saturn, Uranus, and Neptune are tilted relative to their rotation axes. Note that the magnetic fields of Uranus and Neptune

are offset from the centers of the planets and steeply inclined to their rotation axes. Jupiter, Saturn, and Neptune have north magnetic poles on the hemisphere where Earth has its south magnetic pole.

perpendicular to the plane of the ecliptic. The moons were forced into these unusual orbits from whatever their original orbits may have been by the tides they created on the planet. These tides pulled back on the moons and caused them to change orbital direction

until they were over Uranus's equator, where they have been ever since. Five of these satellites, ranging in diameter from 480 to nearly 1600 km (300 to nearly 1000 mi), were known before the *Voyager* mission. *Voyager's* cameras discovered 10 additional satellites,

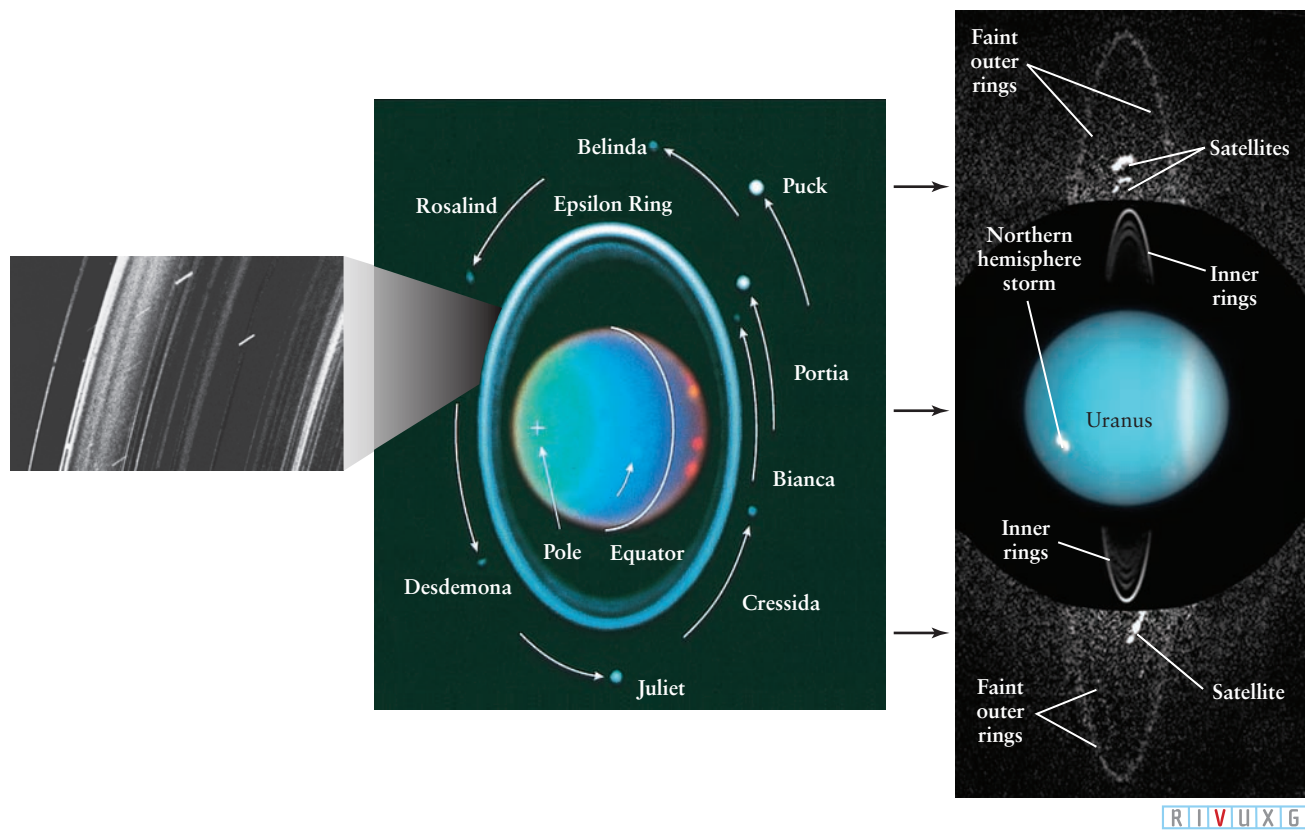


FIGURE 7-38 The Rings and Moons of Uranus (Right) Full-scale image of Uranus and its inner and outer rings. **(Center)** This image of Uranus, its rings, and eight of its moons was taken by the Hubble Space Telescope. **(Left)** Close-up of part of the ring system taken by *Voyager 2* when the spacecraft was

in Uranus's shadow looking back toward the Sun. Numerous fine dust particles between the main rings gleam in the sunlight. The short streaks are star images blurred because of the spacecraft's motion during the exposure. (right: NASA, ESA, and M. Showalter; center: NASA Planetary Photojournal; left: NASA/JPL)

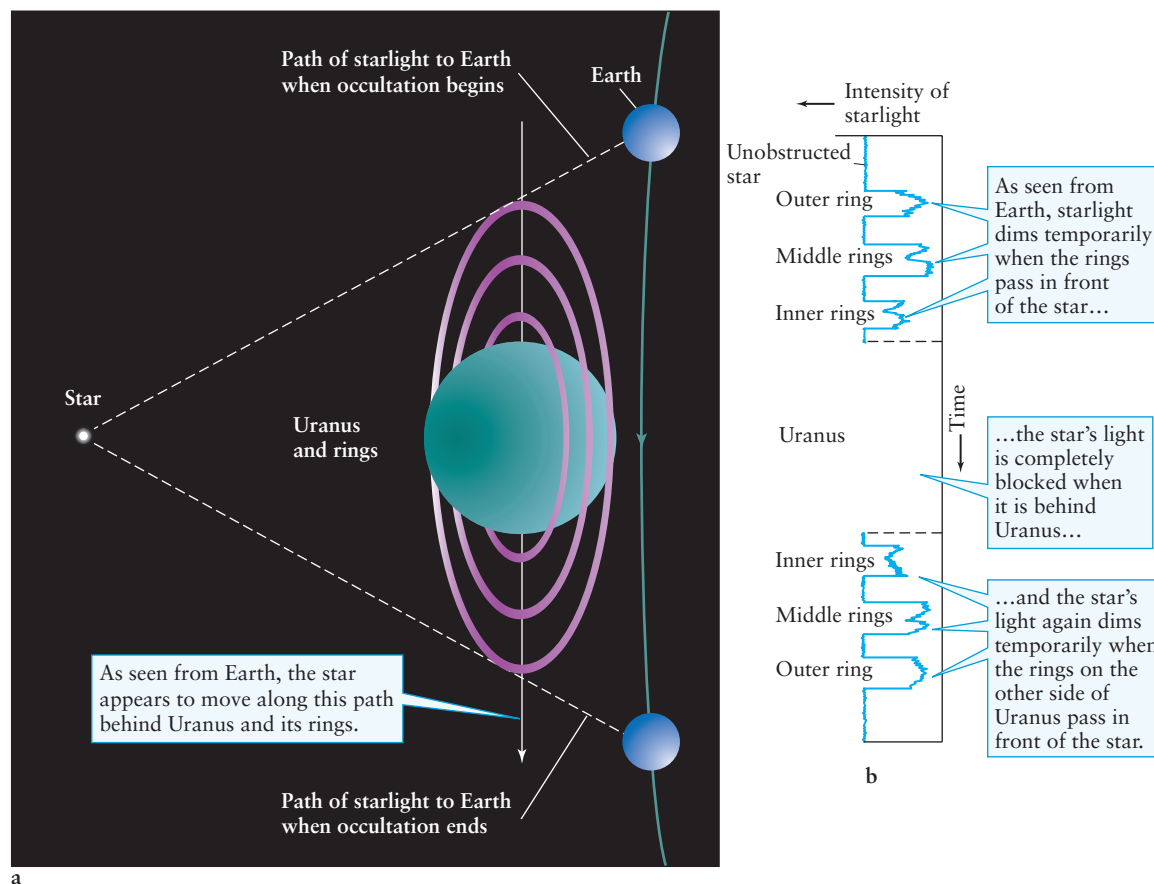


FIGURE 7-39 Discovery of the Rings of Uranus (a) Light from a star is reduced as the rings move in front of it. (b) With sensitive light detectors, astronomers can detect the variation

in light intensity. Such dimming led to the discovery of Uranus's rings. Of course, the star vanishes completely when Uranus occults it.

each smaller than 50 km (31 mi) across. Still others have since been observed from Earth. Several of these tiny, irregularly shaped moons are among the shepherd satellites.

Comparing observations taken 30 years apart, between the 1970s and the first decade of the twenty-first century, reveals that the moons in the vicinity of the rings are changing orbits due to their mutual gravitational influences and the gravitational tugs of the rings. Furthermore, some of the ring debris is being lost and replaced by debris blasted off the moons as a result of meteorite impacts. Observations suggest that the moon Mab, discovered in 2003, is supplying the dust for the outermost ring.

Miranda is the most fascinating and bizarre of Uranus's moons. Unusual wrinkled and banded features cover Miranda's surface (Figure 7-40). Its highly varied terrain suggests that it was once seriously disturbed. Perhaps a shattering impact temporarily broke it into several pieces that then recoalesced, or perhaps severe tidal heating, as we saw on Io, rearranged large pieces of its surface.

Miranda's core originally consisted of dense rock, while its outer layers were mostly ice. If a powerful impact did occur, blocks of debris broke off from Miranda and then drifted back together through mutual gravitational attraction. Recolliding with that moon, they formed a chaotic mix of rock and ice. In this scenario, the landscape we see today on Miranda is the result of huge, dense rocks trying to settle toward the satellite's center, forcing blocks of less dense ice upward toward the surface.

NEPTUNE

Neptune is physically similar to Uranus (review Figures 7-34 and 7-36). Neptune has 17.1 times Earth's mass, 3.88 times Earth's diameter, and a density of 1640 kg/m^3 (2760 lb/yd^3). Its rapid rotation draws its clouds into belts and zones. Unlike on Uranus, however, cloud features can readily be discerned on Neptune. Its whitish, cirruslike clouds (Figure 7-41b) consist of methane ice crystals. As on Uranus, the methane absorbs red light, leaving the planet's belts and zones with a banded, bluish appearance (Figure 7-41). Like the other giant

Focus Question 7-13

Why do Uranus's rings remain in orbit?

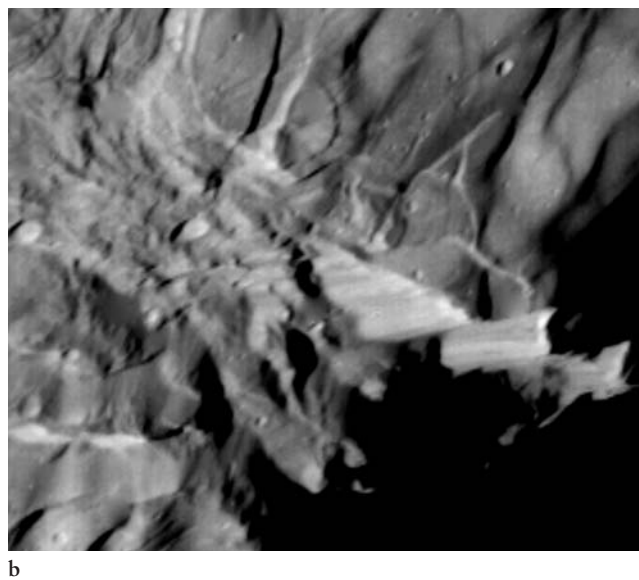
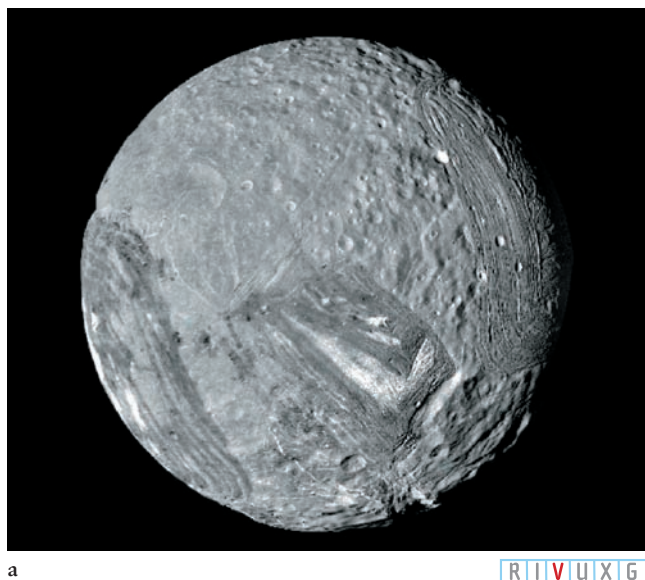


FIGURE 7-40 Miranda (a) The patchwork appearance of Miranda in this mosaic of *Voyager 2* images suggests that this satellite consists of huge chunks of rock and ice that came back together after an ancient, shattering impact by an asteroid or a neighboring Uranian moon. The curious banded features that cover much of Miranda are parallel valleys and ridges that

may have formed as dense, rocky material sank toward the satellite's core. At the very bottom of the image and—where a “bite” seems to have been taken out of the satellite—is a range of enormous cliffs that jut upward as high as 20 km (12 mi), twice the height of Mount Everest. (b) Close-up of cliffs on Miranda. (a: NASA; b: NASA/JPL)

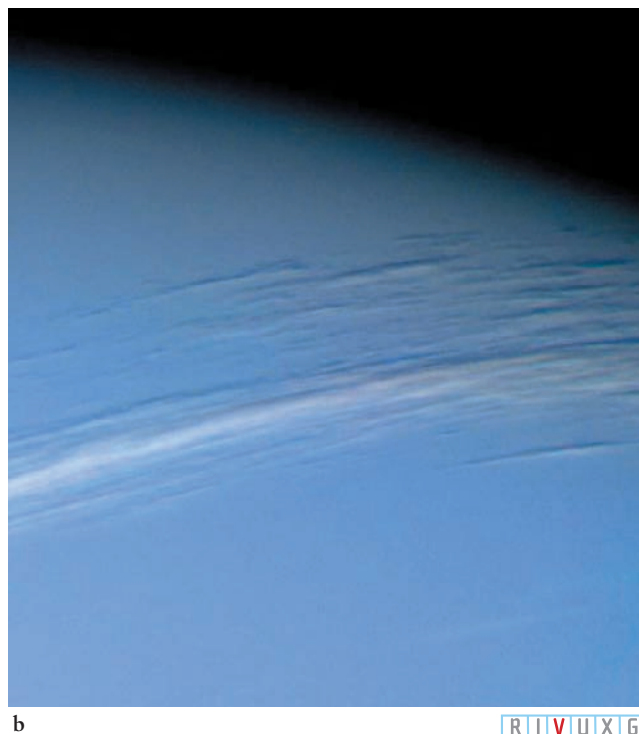
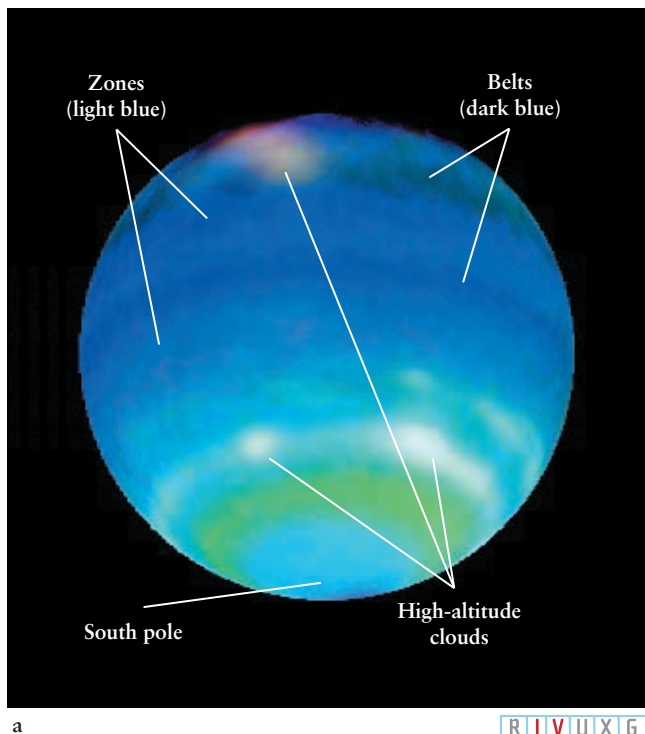


FIGURE 7-41 Neptune's Banded Structure (a) Several Hubble Space Telescope images at different wavelengths were combined to create this enhanced-color view of Neptune. The dark blue and light blue areas are the belts and zones, respectively. The slightly darker belt, running across the middle of the image, lies just south of Neptune's equator. White areas are

high-altitude clouds of methane ice. The very highest clouds are shown in yellow-red, as seen at the very top of the image. The green belt near the south pole is a region where the atmosphere absorbs blue light, probably indicating some differences in chemical composition. (b) Methane clouds above the belts and zones of Neptune. (a: NASA/JPL/STScI; b: NASA/JPL)

planets, the atmosphere of Neptune experiences differential rotation. The winds on Neptune blow at speeds of up to 2000 km/h (1240 mi/h)—among the fastest known in the solar system. Neptune’s rotation axis is tilted nearly 30° from the plane of its orbit around the Sun and Hubble Space Telescope observations of Neptune in 2003 showed seasonal changes near its poles.

7-16 Neptune was discovered because it had to be there

Neptune’s discovery is storied because it illustrates a scientific prediction leading to an expected discovery. Recall from Section 2-8 that, in 1781, the British astronomer William Herschel discovered Uranus. The planet’s position was carefully plotted and, by the 1840s, it was clear that even considering the gravitational effects of all of the known bodies in the solar system, Uranus was not following the path predicted by Newton’s and Kepler’s laws. Either the theories behind these laws were wrong or there had to be another, yet-to-be-discovered body in the solar system pulling on Uranus.

Independent, nearly simultaneous, calculations by an English mathematician, John Adams, and a French astronomer, Urbain Leverrier, predicted the same location for the alleged planet. That planet, Neptune, was located in 1846 by the German astronomer Johann Galle, within a degree or two of where it had to be to have the observed influence on Uranus.

In August 1989, nearly 150 years after Neptune’s discovery, *Voyager 2* arrived at the planet to cap one of NASA’s most ambitious space missions. Scientists were overjoyed at the detailed, close-up pictures and wealth of data about Neptune sent back to Earth by the spacecraft.

At the time *Voyager 2* passed it, a giant storm raged in Neptune’s atmosphere. Called the **Great Dark Spot**, the storm was about half as large as Jupiter’s Great Red Spot. The Great Dark Spot (Figure 7-42) was located at about the same latitude on Neptune and occupied a similar proportion of Neptune’s surface as the Great Red Spot does on Jupiter. Although these similarities suggested that similar mechanisms created the spots, the Hubble Space Telescope in 1994 showed that the Great Dark Spot had disappeared. Then, in April 1995, another spot developed in the opposite hemisphere.

Neptune’s interior is believed to be very similar in composition and structure to that of Uranus: a rocky core surrounded by ammonia and methane-laden water (see Figure 7-36). Neptune’s surface magnetic field is about 40% that of Earth. As with Uranus, Neptune’s magnetic axis (the line connecting its north and south magnetic poles) is tilted sharply from its rotation axis.

In this case, the tilt is 47° . Also like Uranus, Neptune’s magnetic axis does not pass through the center of the planet (see Figure 7-37).

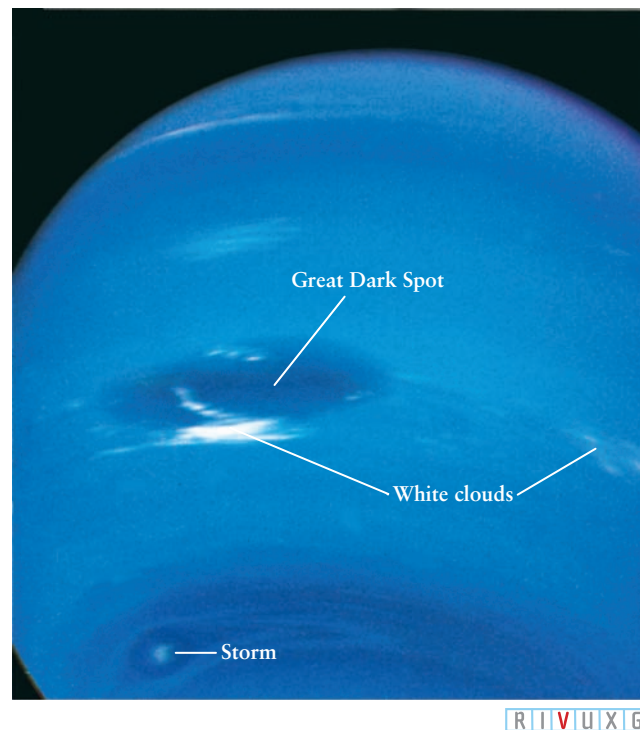


FIGURE 7-42 Neptune This view from *Voyager 2* looks down on the southern hemisphere of Neptune. The Great Dark Spot’s longer dimension at the time was about the same size as Earth’s diameter. It has since vanished. Note the white, wispy methane clouds. (NASA/JPL)

Insight Into Science

Process and Progress Astronomers were able to predict Neptune’s position on the basis of theory before it was observed. Likewise, the volcanoes on Io were also predicted. In other cases in science, theories have correctly predicted results even before the technology existed to verify them. Once confirmed, valid theories lead to other predictions and further scientific understanding.

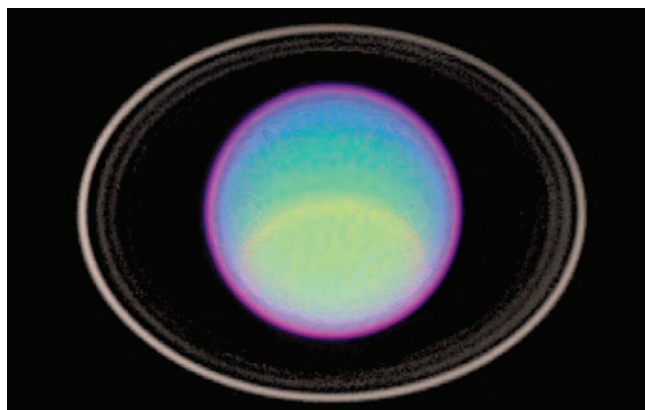
We saw in Section 7-2 that the magnetic fields of Jupiter and Saturn are believed to be generated by the motions of their liquid metallic hydrogen. However, Uranus and Neptune lack this material, and their magnetic fields have a different origin. These fields are believed to exist because molecules such as ammonia, which dissolve in their water layers, lose electrons (become ionized). These ions, moving with the planets’ rotating, fluid interiors, create the same dynamo effect and, hence, also the magnetic field.

7-17 Neptune has rings and captured moons

Like Uranus, Neptune is surrounded by a system of thin, dark rings (Figure 7-43). At these distances from the Sun it is so cold that both planets’ ring particles retained

Focus Question 7-14

The “spots” on Neptune are analogous to what features in Earth’s atmosphere?



R I V U X G

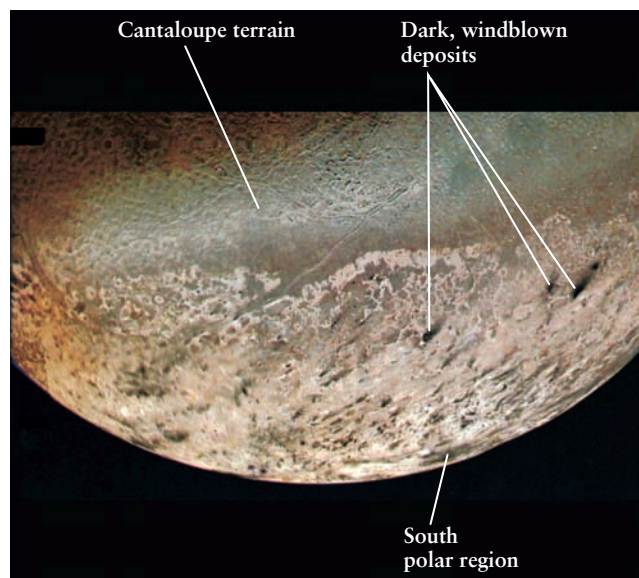
FIGURE 7-43 Neptune's Rings Two main rings are easily seen in this Hubble image. Careful examination also reveals a faint inner ring. A fainter-still sheet of particles, whose outer edge is located between the two main rings, extends inward toward the planet. (Erich Karkoschka (University of Arizona Lunar & Planetary Lab) and NASA)

methane ice. Scientists speculate that eons of radiation damage have converted this methane ice into darkish carbon compounds, thus accounting for the low reflectivity of the rings.

Neptune has 14 known moons. Thirteen have irregular shapes and highly elliptical orbits, which suggest that Neptune captured them. Triton, discovered in 1846, is spherical and was quickly observed to have a nearly circular, retrograde orbit around Neptune. It is difficult to imagine how a planet rotating one way and a satellite revolving the other way could form together. Indeed, only the small outer satellites of Jupiter and Saturn have retrograde orbits, and these bodies are probably captured asteroids. Some scientists have therefore suggested that Triton was also captured 3 or 4 billion years ago by Neptune's gravity. Upon being captured, Triton was most likely in a highly elliptical orbit. However, the tides that the moon creates on Neptune's liquid surface would, in turn, have made Triton's orbit more circular.

Triton's average density of about 2100 kg/m^3 (3540 lb/yd^3) indicates that it is an equal mix of rock and ice. It has a thin nitrogen, methane, and carbon dioxide atmosphere, which changes density with the seasons on that world. When Triton was younger, the tidal force exerted on it by Neptune due to the moon's initially elliptical orbit caused Triton to stretch and flex, providing enough energy to melt much of its interior and obliterate its original surface features, including craters. Triton's south polar region is shown in **Figure 7-44**. Note that very few craters are visible. While cratering continued, so did resurfacing. Calculations indicate that Triton's present surface is only about 100 million years old.

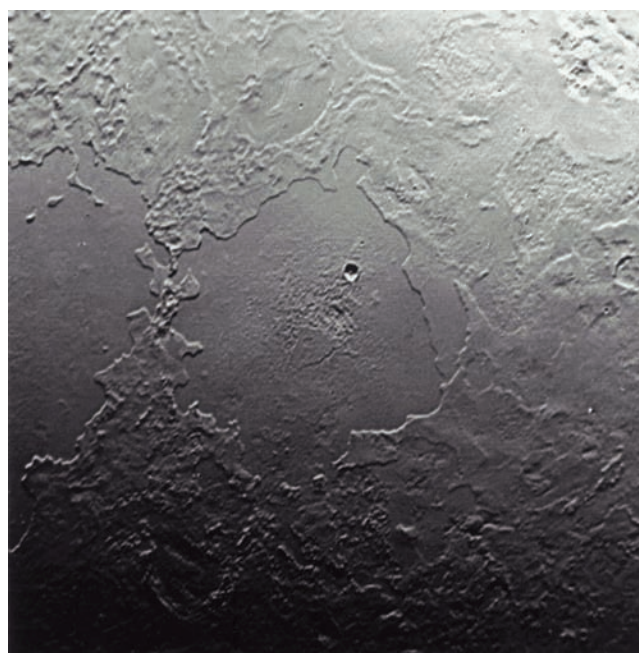
Triton does exhibit some surface features seen on other icy worlds, such as long cracks resembling those on Europa and Ganymede. Other features unique to



R I V U X G

FIGURE 7-44 Triton's South Polar Cap Approximately a dozen high-resolution Voyager 2 images were combined to produce this view of Triton's southern hemisphere. The pinkish polar cap is probably made of nitrogen frost. A notable scarcity of craters suggests that Triton's surface was either melted or flooded by icy lava after the era of bombardment that characterized the early history of the solar system. (NASA/JPL)

Triton are quite puzzling. For example, the top part of Figure 7-44 reveals a wrinkled terrain that resembles the skin of a cantaloupe. Triton also has a few frozen lakes like the one shown in **Figure 7-45**. Some scientists have



R I V U X G

FIGURE 7-45 A Frozen Lake on Triton Scientists think that the feature in the center of this image is a basin filled with water ice. The flooded basin is about 200 km (125 mi) across. (NASA)

speculated that these lakelike features are the calderas of extinct cryovolcanoes. A mixture of methane, ammonia, and water, which can have a melting point far below that of pure water, could have formed a kind of cold lava on Triton.

Voyager instruments measured a surface temperature of 36 K (−395°F), making Triton the coldest world that our probes have ever visited. Nevertheless, *Voyager* cameras did glimpse two towering plumes of gas extending up to 8 km (5 mi) above the satellite's surface. These are apparently jets of nitrogen gas warmed by interior radioactive decay and escaping through vents or fissures.

Triton continues to create tides on Neptune. Whereas the tides on Earth cause our Moon to spiral outward, the tides on Neptune cause Triton (in its retrograde orbit) to spiral inward. Within the next quarter of a billion years, Triton will reach the Roche limit, at which time pieces of Triton will then literally float into space until the entire moon is demolished! By destroying Triton, Neptune will create a new ring system that will be much more substantial than its present one (Figure 7-46).

Focus Question 7-15

Is our Moon inside or outside Earth's Roche limit?

COMPARATIVE PLANETOLOGY OF THE OUTER PLANETS

Now that we have examined the individual outer planets, it is instructive to compare their various properties. **Table 7-1, The Outer Planets: A Comparison**, summarizes much of this material.

Size and mass The four giant planets, Jupiter, Saturn, Uranus, and Neptune, are all much larger and more massive than Earth (that is, roughly 4 to 11 times Earth's diameter and 14 to 318 times Earth's mass).

Atmospheres and rotations The four outer planets have thick hydrogen- and helium-rich atmospheres that are permanently and completely covered with clouds. All of the giant planets rotate rapidly compared to Earth. Their sidereal rotation rates range from about 10 hours to about 17¼ hours. This rapid rotation draws their clouds into parallel bands, called *belts* and *zones*. Earth's slower rotation allows the clouds and winds here to roam over a much greater range of latitudes than on the giant planets.

Interiors All four giant planets have terrestrial bodies as their cores. That is where their internal similarities to Earth end. The cores of Jupiter and Saturn are surrounded by thin shells of “ices,” above which lie thick liquid metallic hydrogen layers, which, in turn, are surrounded by layers of normal hydrogen and helium. Calculations suggest that Jupiter's core is being dissolved by the interaction of its iron with its surrounding liquid metallic hydrogen. In contrast, the terrestrial cores of Uranus and Neptune are surrounded by large amounts of water, which, in turn, are surrounded by normal liquid hydrogen and helium.

Magnetic fields and rings The four giant planets all have magnetic fields that store particles from the solar wind, similar to Earth's Van Allen belts. Like Earth's magnetic field, those of Jupiter, Uranus, and Neptune are tilted relative to their rotation axes. Saturn's magnetic field is along its rotation axis. The four

giants all have rings. Saturn's rings are the most massive and distinctive, followed by those of Uranus, Jupiter, and Neptune.

Moons Among them, the giant planets have at least 170 moons. Each of the giants has between one and a few spherical moons, like our Moon. However, most of the

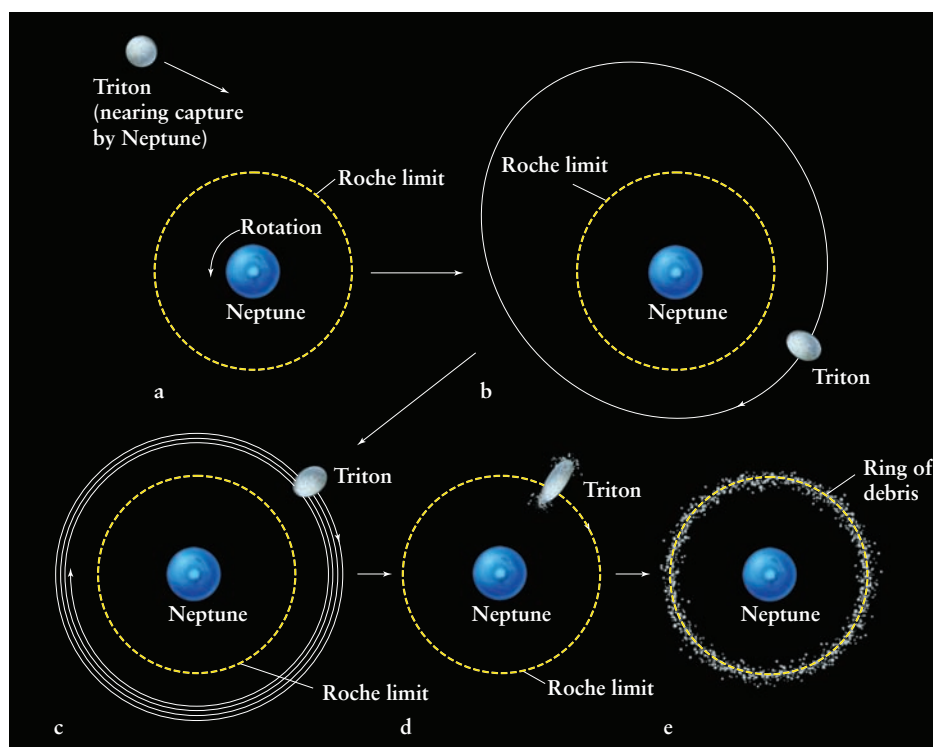
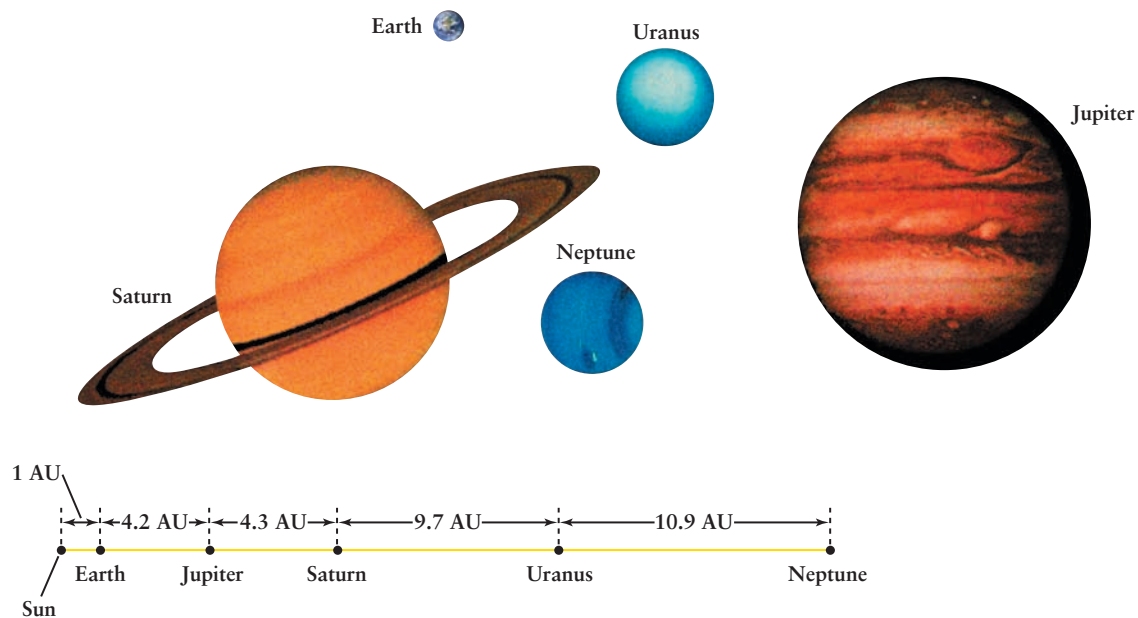


FIGURE 7-46 The Capture and Destruction of Triton This series of drawings depicts how (a) Triton was captured by Neptune in a retrograde orbit. (b) The tides that Triton then created on the planet caused that moon's orbit to become quite circular and (c) to spiral inward. (d) It will eventually reach Neptune's Roche limit and (e) be pulled apart to form a ring. (Stephen P. Meszaros/NASA, M. Buie, K. Horne, and D. Tholen)

TABLE 7-1 THE OUTER PLANETS: A COMPARISON

	Interior	Surface	Rings	Atmosphere	Magnetic Field*
Jupiter	Terrestrial core, liquid “ice” shell, liquid metallic hydrogen shell, liquid hydrogen and helium mantle	No solid surface, atmosphere gradually thickens to liquid state, belt and zone structure, hurricane-like features	Yes	Primarily H, He	$19,000 \times$ Earth’s total field; at its cloud layer, $14 \times$ Earth’s surface field
Saturn	Similar to Jupiter, with bigger terrestrial core and less metallic hydrogen	No solid surface, less distinct belt and zone structure than Jupiter	Yes	Primarily H, He	$570 \times$ Earth’s total field; at its cloud layer, $\frac{2}{3} \times$ Earth’s surface field
Uranus	Terrestrial core, liquid water shell, normal liquid hydrogen and helium mantle	No solid surface, weak belt and zone system, hurricane-like features, color from methane absorption of red, orange, yellow	Yes	Primarily H, He, some CH_4	$50 \times$ Earth’s total field; at its cloud layer, $0.73 \times$ Earth’s surface field
Neptune	Similar to Uranus	Similar to Uranus	Yes	Primarily H, He, some CH_4	$35 \times$ Earth’s total field; at its cloud layer, $0.4 \times$ Earth’s surface field



For detailed numerical comparisons between planets, see Appendix C and Appendix D.

*To see the orientations of these magnetic fields relative to the rotation axes of the planets, see Figure 7-37.

moons in the solar system are much smaller, irregularly shaped bodies. These objects are most likely captured space debris. Unlike our Moon, all of the satellites of the giants are at least a few hundred thousand times less massive than their planets. Indeed, most moons are mil-

lions or billions of times less massive than their planets. Recall that our Moon is only 81 times less massive than Earth.

Focus Question 7-16

How do the numbers and types of moons found orbiting the giant planets compare to those that orbit the terrestrial planets?

SUMMARY OF KEY IDEAS

Jupiter and Saturn

- Jupiter is by far the largest and most massive planet in the solar system.
- Jupiter and Saturn have rocky cores surrounded by a thin shell of ices, then a thick shell of liquid metallic hydrogen and an outer layer of ordinary liquid hydrogen and helium. Jupiter's core may be dissolving. Both planets have an overall chemical composition very similar to that of the Sun.
- The visible features of Jupiter exist in the outermost 100 km (62 mi) of its atmosphere. Saturn has similar features, but they are much fainter.
- Three cloud layers exist in the upper atmospheres of both Jupiter and Saturn. Because Saturn's cloud layers extend through a greater range of altitudes, the colors of the Saturnian atmosphere appear muted.
- The colored ovals visible in the Jovian atmosphere are gigantic storms, some of which (such as the Great Red Spot) are stable and persist for years or even centuries.
- Jupiter and Saturn have strong magnetic fields created by electric currents in their metallic hydrogen layers.
- Four large satellites orbit Jupiter. The two inner Galilean moons, Io and Europa, are roughly the same size as our Moon. The two outer moons, Ganymede and Callisto, are approximately the size of Mercury.
- Io is covered with a colorful layer of sulfur compounds deposited by frequent explosive eruptions from volcanic vents. Europa is covered with a smooth layer of frozen water crisscrossed by an intricate pattern of long cracks.
- The heavily cratered surface of Ganymede is composed of frozen water with large polygons of dark, ancient crust separated by regions of heavily grooved, lighter-colored, younger terrain. Callisto has a heavily cratered ancient crust of frozen water.
- Both Jupiter and Saturn have rings. Saturn is circled by a system of thin, broad rings lying in the plane of the planet's equator. Each major ring is composed of a great many narrow ringlets that consist of numerous fragments of ice and ice-coated rock. Jupiter has a much less substantial ring system.
- Saturn's moon Titan has a thick atmosphere of nitrogen, methane, and other gases, as well as lakes of methane and ethane.
- Saturn's moon Enceladus has areas with very different surface features: an older, heavily cratered region and a newer, nearly crater-free surface created by tectonic activity.

Uranus and Neptune

- Uranus and Neptune are quite similar to each other in appearance, mass, size, and chemical composition. Each

has a rocky core surrounded by a thick, watery mantle topped by a layer rich in normal hydrogen and helium; the axes of their magnetic fields are steeply inclined to their axes of rotation; and both planets are surrounded by systems of thin, dark rings.

- Uranus is unique in that its axis of rotation lies near the plane of its orbit, producing greatly exaggerated seasons on the planet.
- Uranus has five moderate-sized satellites, the most bizarre of which is Miranda.
- Triton, the largest satellite of Neptune, is an icy world with a tenuous nitrogen atmosphere. Triton moves in a retrograde orbit that suggests it was captured into orbit by Neptune's gravity. It is spiraling down toward Neptune and will eventually break up and form a ring system.

WHAT DID YOU THINK?

1 *Is Jupiter a "failed star"? Why or why not?*

No. Jupiter has 75 times too little mass to shine as a star.

2 *What is Jupiter's Great Red Spot?* The Great Red Spot is a long-lived, oval cloud circulation, similar to a hurricane or typhoon on Earth.

3 *Does Jupiter have continents and oceans?* No. Jupiter is surrounded by a thick atmosphere primarily composed of hydrogen and helium that gradually becomes liquid as you move inward. The only solid matter in Jupiter is its core.

4 *Is Saturn the only planet with rings?* No. All four giant planets (Jupiter, Saturn, Uranus, and Neptune) have rings.

5 *Are the rings of Saturn solid ribbons?* No. Saturn's rings are all composed of thin, closely spaced ringlets consisting of particles of ice and ice-coated rocks. If they were solid ribbons, Saturn's gravitational tidal force would tear them apart.

Review Questions

- Which is the most massive planet in the solar system?
 - Earth
 - Neptune
 - Saturn
 - Jupiter
 - Mercury
- Which of the following planets does not have rings? Choose only one.
 - Mars
 - Uranus
 - Neptune
 - Saturn
 - Jupiter

3. Which is the least massive planet in the solar system?

- a. Mercury
- b. Mars
- c. Uranus
- d. Jupiter
- e. Venus

4. Which planet is presently known to have the most moons?

- a. Mars
- b. Saturn
- c. Uranus
- d. Jupiter
- e. Neptune

5. Describe the appearance of Jupiter's atmosphere. Which features are long-lived and which are relatively fleeting?

6. To test your knowledge of Jupiter's belt and zone structure, do Interactive Exercise 7.1 on the assigned Web site. You can print out your results, if required.

7. What causes the belts and zones in Jupiter's atmosphere?

8. What is liquid metallic hydrogen? Which planets contain this substance? What produces this form of hydrogen?

9. Compare and contrast the surface features of the four Galilean satellites, discussing their geologic activity and their evolution.

10. What energy source powers Io's volcanoes?

11. Why are numerous impact craters found on Ganymede and Callisto but not on Io or Europa?

12. Describe the structure of Saturn's rings. What are they made of?

13. Why do features in Saturn's atmosphere appear to be much fainter and more "washed out" than comparable features in Jupiter's atmosphere?

14. Explain how shepherd satellites affect some planetary rings. Is "shepherd satellite" an appropriate term for these objects? Explain your answer.

15. Describe Titan's atmosphere. What effect does sunlight have on it?

16. Describe Titan's surface.

17. Describe the seasons on Uranus. Why are the Uranian seasons different from those on any other planet?

18. Briefly describe the evidence that supports the idea that Uranus was struck by a large planetlike object several billion years ago.

19. Why are Uranus and Neptune distinctly bluer than Jupiter and Saturn?

20. Compare the ring systems of Saturn and Uranus. Why were Uranus's rings unnoticed until the 1970s?

21. How do the orientations of Uranus's and Neptune's magnetic axes differ from those of the other planets?

22. Explain why Triton will never collide with Neptune, even though Triton is spiraling toward that planet.

Advanced Questions

23. Explain how the tides on planets generated by captured moons force the moons into more circular orbits around the equators of their planets.

24. Long before the *Voyager* flybys, Earth-based astronomers reported that Io appeared brighter than usual for a few hours after emerging from Jupiter's shadow. Explain this brief brightening of Io.

25. Compare and contrast Valhalla on Callisto with Caloris Basin on Mercury.

26. As seen by Earth-based observers, the intervals between successive edge-on presentations of Saturn's rings alternate between 13 years 9 months and 15 years 9 months. Why are these two intervals not equal?

27. Compare and contrast the internal structures of Jupiter and Saturn with the internal structures of Uranus and Neptune. Can you propose an explanation for why the differences between these two pairs of planets occurred?

28. Neptune has more mass than Uranus, but Uranus is the larger of the two planets. Reconcile these two facts.

Discussion Questions

29. Suppose that you were planning a mission to Jupiter, using an airplanelike vehicle that would spend many days, even months, flying through the Jovian clouds. What observations, measurements, and analyses should this aircraft make? What dangers might it encounter, and what design problems would you have to overcome?

30. Discuss why astronomers believe that Europa, Ganymede, and Callisto may harbor some sort of marine life. Why do they not expect any life on the surfaces of these worlds?

31. Suppose you were planning separate missions to each of Jupiter's Galilean moons. What questions would you want these missions to answer, and what kinds of data would you want your spacecraft to send back? Given the different environments on the four satellites, how would the designs of the four spacecraft differ?

32. NASA and the Jet Propulsion Laboratory have tentative plans to place spacecraft in orbit about Uranus and Neptune in this century. What kinds of data should be collected, and what questions would you like to see answered by these missions?

Got It?

33. How would descending into Jupiter differ from descending from space onto Earth?
34. Why does the moon Io have active volcanoes while our Moon does not?
35. Is the Cassini division in Saturn's rings empty? Justify your answer.
36. Jupiter's Great Red Spot is most accurately described as which of the following?
 - a. an impact site
 - b. a dormant volcano
 - c. an active volcano
 - d. a hurricane
 - e. a tornado



Comet Hale-Bopp. (Walter Pacholka, Astropics/Science Source)

CHAPTER 8

Vagabonds of the Solar System

WHAT DO YOU THINK?

- 1 Are the asteroids a former planet that was somehow destroyed? Why or why not?
- 2 How far apart are the asteroids on average?
- 3 How are comet tails formed? Of what are they made?
- 4 In which directions do a comet's tails point?
- 5 What is a shooting star?

Answers to these questions appear in the text beside the corresponding numbers in the margin and at the end of the chapter.

Now that we have explored the properties of the planets and their moons in Chapters 6 and 7, we finish studying the objects orbiting the Sun in this chapter. In Section 4-8 we saw that astronomers have several overlapping ways of classifying the bodies orbiting the Sun. To briefly review, recall that a **planet** is a celestial body that (1) is in orbit around the Sun, (2) has sufficient mass for its self-gravity to pull itself into a nearly spherical shape (although rotation causes many planets to be wider at their equators than at their poles), and (3) also has enough mass for its gravitational force to clear the neighborhood around its orbit. A **dwarf planet** satisfies conditions (1) and (2), but it does not have enough gravity to clear its orbital neighborhood of debris. (Note that the dwarf planet Haumea is egg-shaped, rather than being uniformly wider at its equator, like Earth. This is apparently due to Haumea's exceptionally rapid rotation.) All objects in the solar system that are not planets, dwarf planets, or moons are called **small solar system bodies** (SSSBs).

Focus Question 8-1

What objects are classified today as planets? What objects are classified as dwarf planets?

So far, Pluto, Ceres (also an asteroid), Eris (named after a Greek goddess), Makemake (pronounced *mah-key mah-key* and named after a god of the Rapa Nui people of Easter Island), and Haumea (pronounced *how-may-uh* and named after a goddess of Hawaii) have earned the title of dwarf planet. The latter three bodies were discovered in 2005.

Dozens of other solar system objects are being evaluated for inclusion in this class. Bear in mind that as the properties of more solar system objects are pinned down, the definitions of dwarf planets and SSSBs may be refined.

In this chapter you will discover

- the properties of dwarf planets and small solar system bodies
- asteroids and meteoroids—pieces of rock and metal debris orbiting the Sun
- comets—objects orbiting the Sun that contain large amounts of ice and rocky debris
- meteors—space debris that falls through Earth's atmosphere
- meteorites—space debris that lands on Earth
- the asteroid belt and the Kuiper belt, both filled with a variety of debris orbiting the Sun, including bound pairs of objects
- the impacts from space 250 million and 65 million years ago that caused mass extinctions of life on Earth
- wayward asteroids that could again threaten life on Earth

DWARF PLANETS

8-1 Pluto and its moon, Charon, are about the same size

Pluto (Figure 8-1) was discovered in 1930 by Clyde Tombaugh, who found it by examining photographic plates showing that it moved among the background

stars from night to night (Figure 8-2). Allowing for our improved technology using computers and charge-coupled devices (CCDs) to take the images, this discovery technique is still standard for finding other dwarf planets and small solar system bodies.

Unlike the planets, Pluto's orbital eccentricity of $e = 0.25$ (see Section 2-5) is so great that it is sometimes closer to the Sun than Neptune, as it was from 1979

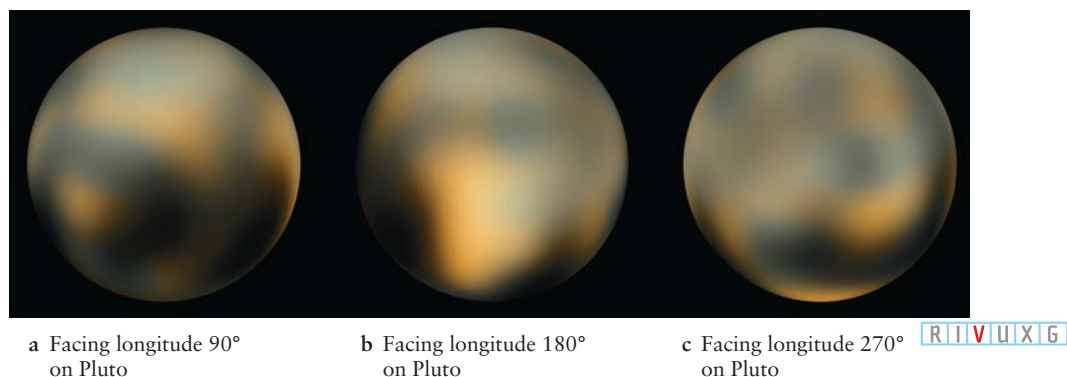


FIGURE 8-1 Pluto These three Hubble Space Telescope images of Pluto show little detail but indicate that the major features of Pluto's surface each cover large amounts of its area.

Comparing these observations to previous ones reveals that the surface changes in color and brightness seasonally. (NASA, ESA, and M. Buie [Southwest Research Institute])

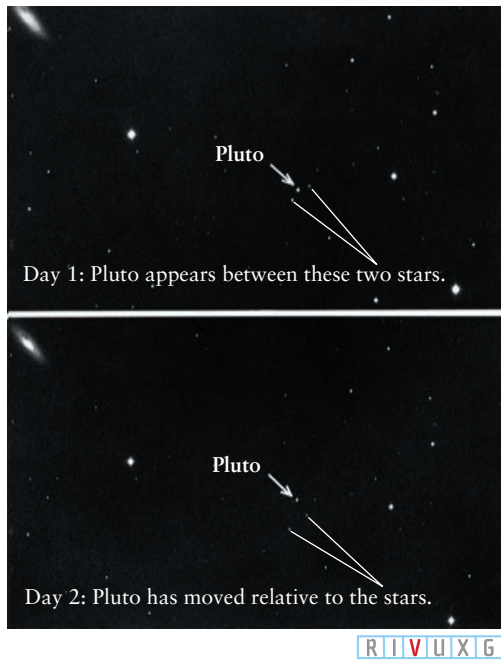


FIGURE 8-2 Discovery of Pluto Pluto was discovered in 1930 by searching for a dim, starlike object that slowly moved against the background stars. These two photographs were taken 1 day apart. (Lick Observatory/Science Source)

to 1999. Pluto is now farther away from the Sun than Neptune and will continue to be so until April 5, 2231 (Figure 8-3). Pluto's orbital inclination, the tilt of its orbit with respect to the plane of the ecliptic, is just over 17° , far greater than that of any planet (Figure 8-3c).

Little was known about Pluto for half a century until astronomers noticed that its image sometimes appears oblong (Figure 8-4). This observation led to the discovery in 1978 of Pluto's largest moon, Charon (pronounced *KER-en*, after the mythical boatman who ferried souls across the River Styx to Hades, the domain ruled by the Roman god Pluto). The mass of an object in space can only be determined directly from its gravitational attraction on another body. We know the mass of the Sun because we can determine the length of Earth's year. We know the mass of Mars because we can determine the periods of its moons Phobos and Deimos. The mass of Pluto, $2.2 \times 10^{-3} M_\oplus$, was not known until the orbit of Charon was accurately determined in the early 1980s.

From 1985 through 1990, the orbit of Charon was oriented so that Earth-based observers could watch it

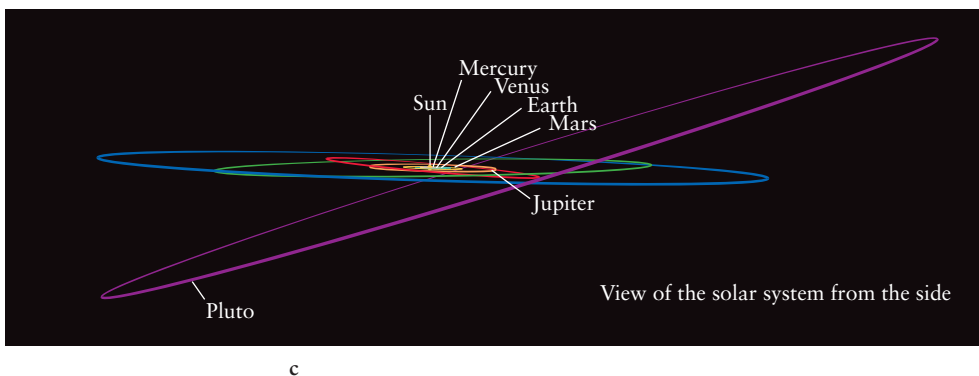
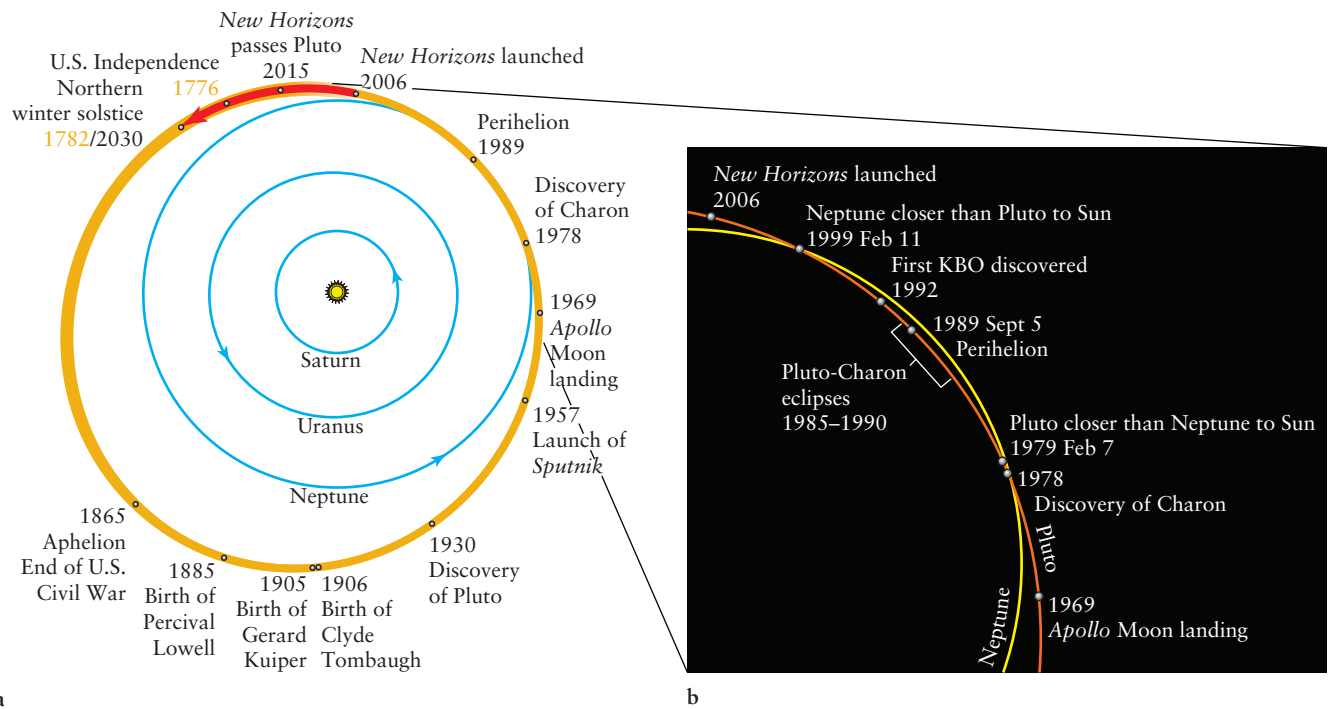


FIGURE 8-3 Orbit of Pluto (a) The high-eccentricity orbit of dwarf planet (and KBO) Pluto stands out compared to the orbits of the outer three planets. Notice how many significant events occurred on Earth during Pluto's present orbit of the Sun. (b) Details of Pluto's passage inside the orbit of Neptune. The two bodies will never collide. (c) A nearly edge-on view of the ecliptic and Pluto's orbit compared to it.

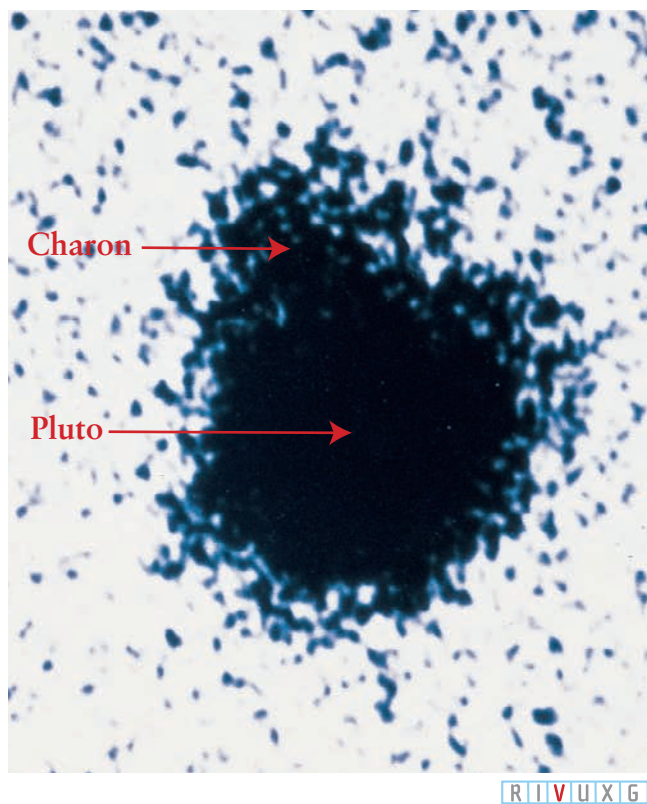


FIGURE 8-4 Discovery of Charon Long ignored as just a defect in the photographic emulsion, the bump on the upper left of this image of Pluto led astronomer James Christy to discover the moon Charon. (U.S. Naval Observatory/NASA)

eclipse Pluto. Astronomers used observations of these eclipses to determine that Pluto is only twice as wide as its satellite: Pluto's diameter is 2380 km (1480 mi) and Charon's is 1190 km. (740 mi). Their similarity in size has led to a debate as to whether they should be classified as a binary dwarf-planet system.

The average distance between Charon and Pluto is less than one-twentieth the distance between Earth and our Moon. Furthermore, both bodies always keep the same side facing each other. That is to say, they both have synchronous rotation with respect to each

other. As seen from the satellite-facing side of Pluto, Charon neither rises nor sets, but instead hovers in the sky, perpetually suspended in the same place above the horizon.

The exceptional similarities between Pluto and Charon suggest that this binary system may have formed when Pluto collided with a body of similar size. Perhaps chunks of matter were stripped from this second body, leaving behind a mass, now called Charon, vulnerable to capture by Pluto's gravity. Alternatively, perhaps Pluto's gravity captured Charon into orbit during a close encounter between the two worlds. For either of these two scenarios to be feasible, many Pluto-sized objects must have existed in the outer regions of the young solar system. One astronomer estimates that there had to be at least a thousand Pluto-sized bodies for a collision or close encounter between two of them to have occurred at least once since the solar system formed 4.6 billion years ago.

In 2005, the Hubble Space Telescope detected two other moons orbiting Pluto (**Figure 8-5**). Called Nix (named after the goddess of the night and mother of Charon) and Hydra (named after the nine-headed serpent living at the gates of Hades), these moons are both much smaller than Charon. The fourth and fifth known moons, temporarily labeled Kerberos and Styx, were discovered in 2011 and 2012, respectively.

Like Neptune's moon Triton, Pluto and Charon are probably both composed of nearly equal amounts of rock and ice. The surface of Pluto (see **Figure 8-1**) has 12 distinct regions—more large-scale features than any object in the solar system other than Earth. Such features on Earth include oceans and continents, while on Pluto they are likely to include regions dominated by rock, regions dominated by ice, major impact regions, and mountain ranges. It is too early to say anything about the physical properties of Nix, Hydra, Kerberos, or Styx.

Studies of Pluto's surface show that different regions vary in temperature from one another by up to 25 K. Furthermore, its surface features are changing in color and brightness, possibly as a result of seasonal effects.

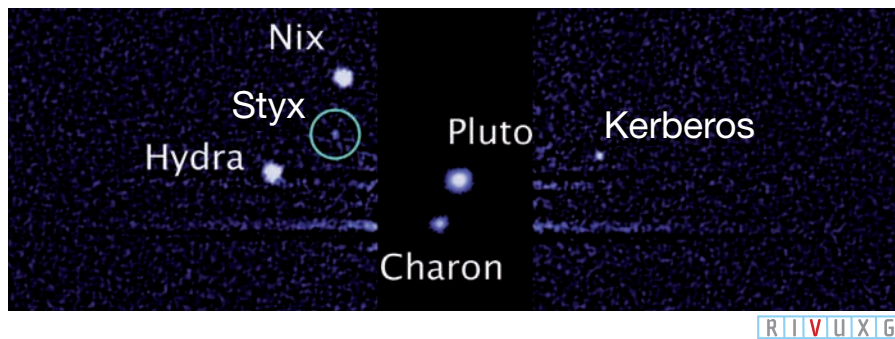


FIGURE 8-5 Pluto and Its Five Known Moons Observations by the Hubble Space Telescope in 2005 revealed two intermediate-sized moons, each about 5000 times dimmer than Pluto. Named Nix and Hydra, they are each about 75 km (50 mi) across and

between 2 and 3 times farther from Pluto than its moon Charon. Kerberos and Styx are each about 25 km (15 mi) in diameter. (NASA; ESA; M. Showalter, SETI Institute)

Specifically, it is getting redder and its northern hemisphere is getting brighter. Its spectrum shows that its surface contains frozen nitrogen, methane, and carbon monoxide. The orange tinge may be due to the presence of carbon-rich compounds on its surface, the result of methane there being chemically altered by the Sun's ultraviolet radiation. The brightest regions are believed to be covered with frozen carbon monoxide, while the darker regions probably contain rocky debris. The dwarf planet is also observed to have a very thin atmosphere of nitrogen and carbon monoxide. In contrast, Charon's surface appears to be covered predominantly with water ice. So intriguing, unusual, and unexplored are Pluto and its companions that NASA launched the *New Horizons* spacecraft in 2006 to visit them (scheduled to arrive there in July 2015) and then to continue deeper into the Kuiper belt. Kuiper belt objects (KBOs) were first introduced in Section 4-5.

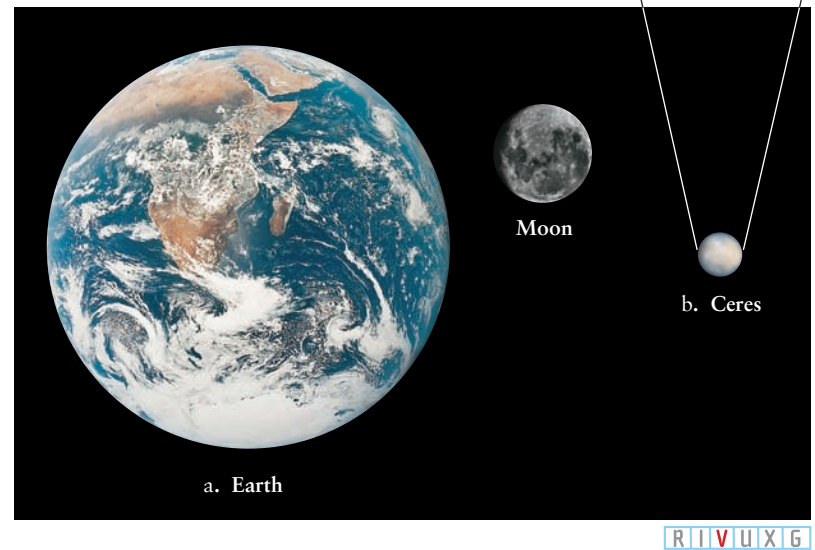
8-2 Ceres is a dwarf planet in the asteroid belt, while Pluto, Eris, Haumea, and Makemake are trans-Neptunian objects as well as dwarf planets

On New Year's Day 1801, the Sicilian astronomer Giuseppe Piazzi was carefully mapping faint stars in the constellation Taurus. He noticed a dim, previously uncharted "star" that shifted its position slightly over the next several nights. Uranus had been discovered in this way by William Herschel just 20 years before. Later that year, the orbit of Piazzi's object was determined to lie between Mars and Jupiter. At his request, it was named Ceres (pronounced *See-reez*), after the patron goddess of Sicily. Based on erroneous observations of its size, Ceres was initially catalogued as a planet. In 1855, further observations revealed that all of the then-known planets—Mercury, Venus, Earth, Mars, Jupiter, Saturn, and Uranus—were much larger than Ceres. We now know that Piazzi was the first astronomer to discover an asteroid.

Ceres is spherical like the planets (Figure 8-6), but its diameter is a scant 940 km (585 mi), only one-quarter the diameter of our Moon. In 2006, Ceres was classified as a dwarf planet. Recent calculations suggest that, like the planets, Ceres is differentiated, containing a rocky core and clad in an icy outer layer. In 2013, plumes of water vapor were observed around Ceres. The mechanism for this water's release is still under investigation.

Eris was first observed in 2003, before the present definitions of planets, dwarf planets, and SSSBs were established. This body is presently located 97 AU from the Sun, 3 times farther out than Pluto. All bodies orbiting the Sun beyond Neptune are called *trans-Neptunian objects* or TNOs. Pluto, Eris, Makemake, and Haumea are dwarf planets and TNOs. Furthermore, its 557-year orbit is so eccentric that for part of that time Eris is closer

FIGURE 8-6 Comparison of Ceres with the Moon and Earth Ceres, the Moon, and Earth are shown here to scale. Dwarf planet Ceres is the largest asteroid. This image of Ceres suggests that it has regions of ice and rock on its surface. The asteroid will be visited by the *Dawn* spacecraft in 2015. (NASA)



to the Sun than Pluto (Figure 8-7a). Eris has a diameter of about 2400 km/1490 mi (Pluto is 2300 km/1430 mi across) and is very bright, with an albedo of about 0.86. Observations suggest that one-quarter of this world is water ice, a greater volume than all of the freshwater on Earth. A moon orbiting Eris, named Dysnomia (after the daughter of the Greek goddess Eris), was discovered in 2005 (Figure 8-7b). The moon is about 300 km (185 mi) across.

Focus Question 8-2

Where is Eris located in the solar system?

Haumea, discovered in 2005, rotates once every 3 hours 55 minutes. In other words, a day on Haumea is less than 4 hours long. Because of that rapid rotation, this world has distorted itself so that it looks like a watermelon. Haumea is a dwarf planet even though its rapid rotation deforms it from spherical, which calculations show it would be if it were not spinning. Haumea has two known moons, H'iaka (about 310 km across) and Namaka (roughly 170 km across). They are named after Haumea's mythological daughters. Spectra of both moons reveal that unlike the rest of the bodies observed in the outer solar system, their surfaces are made of almost pure water ice. In contrast, most objects have surfaces of rock and ice, where the "ice" often contains molecules other than water.

Makemake, also discovered in 2005, rotates once every 7 hours 45 minutes. It is about three-quarters the size of Pluto and it orbits the Sun once every 310 years. Unlike the other Kuiper belt dwarf planets, Makemake has no known moons.

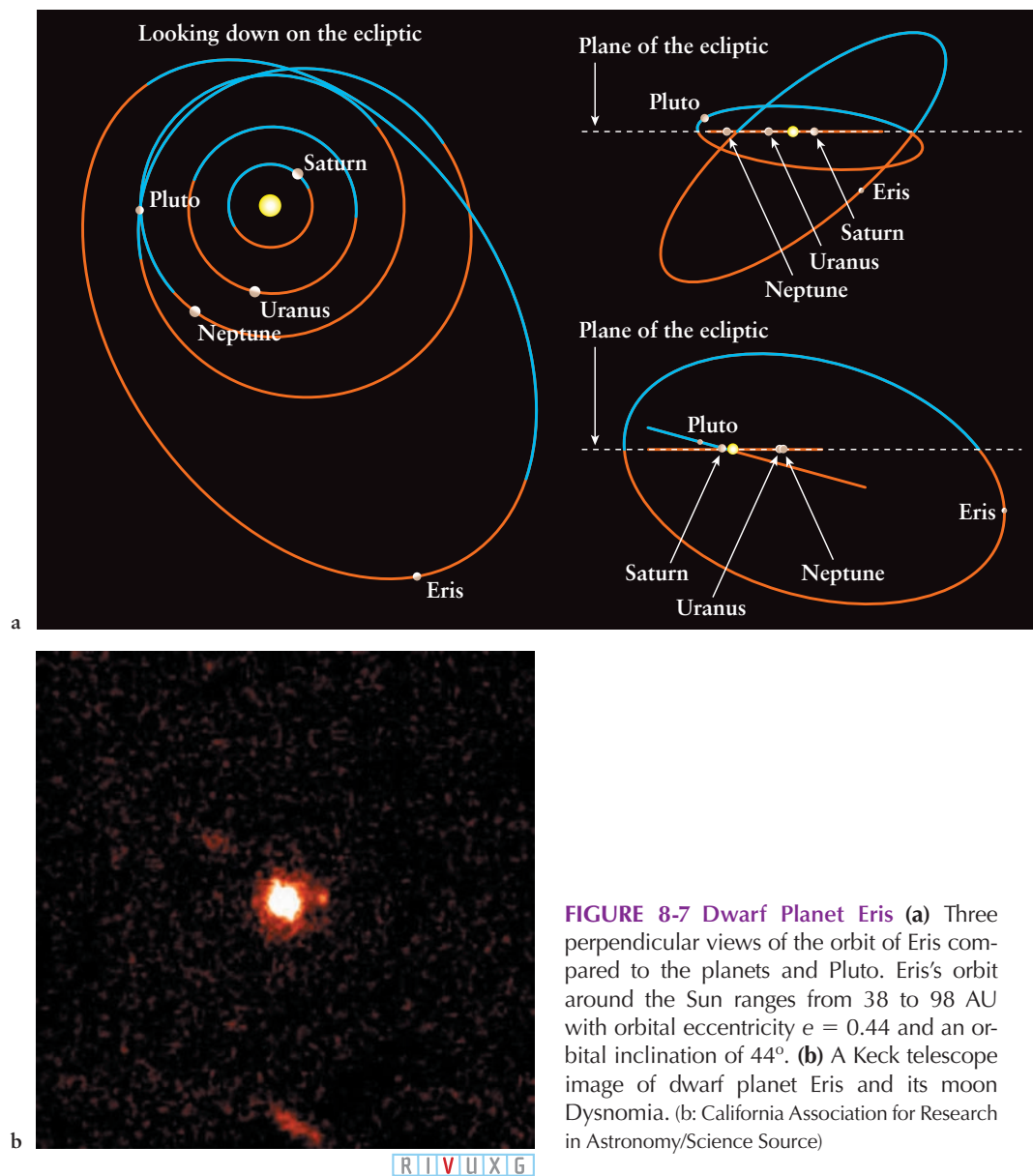


FIGURE 8-7 Dwarf Planet Eris (a) Three perpendicular views of the orbit of Eris compared to the planets and Pluto. Eris's orbit around the Sun ranges from 38 to 98 AU with orbital eccentricity $e = 0.44$ and an orbital inclination of 44° . (b) A Keck telescope image of dwarf planet Eris and its moon Dysnomia. (b: California Association for Research in Astronomy/Science Source)

While astronomers decide which other bodies will be designated dwarf planets, we turn to the small solar system bodies, which include the rest of the asteroids, all the comets, and the smallest rubble of all, the meteoroids.

SMALL SOLAR SYSTEM BODIES

We know that the solar system formed from a rotating disk of gas and dust. Matter in the disk with too much angular momentum to fall onto the protosun coalesced at varying orbital distances into planetesimals (see Section 4-2 for more on planetesimals). Many of these chunks of rock and metal eventually collided, forming the planets, dwarf planets, and larger moons of the solar system, as well as myriad pieces of dust, rock, and boulder-sized debris. Others were captured whole by various planets as

small, irregularly shaped moons, like Phobos and Deimos in orbit around Mars. Many planetesimals still orbit the Sun today in splendid isolation. These are the asteroids, sometimes called *minor planets* (not to be confused with dwarf planets).

ASTEROIDS

8-3 Most asteroids orbit the Sun between Mars and Jupiter

In 1802, less than two years after the asteroid Ceres (now also a dwarf planet) was discovered, the German astronomer Heinrich Olbers observed another faint, starlike object that moved against the background stars. He called it Pallas, after the Greek goddess of wisdom. Like Ceres, Pallas orbits the Sun in a nearly circular path

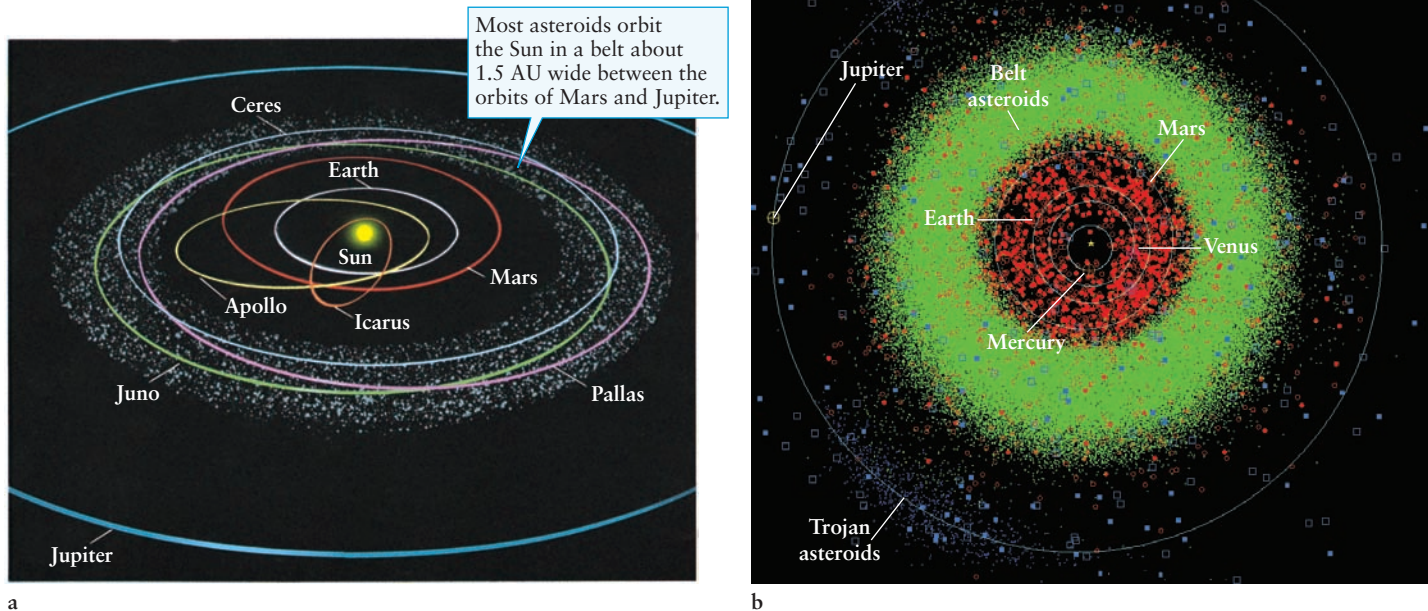


FIGURE 8-8 Asteroid Orbits (a) The orbits of belt asteroids Ceres, Pallas, and Juno are indicated to scale in this diagram. Some asteroids that are not in the asteroid belt, such as Apollo and Icarus, have highly eccentric paths that cross Earth's orbit. Other non-belt asteroids, called the Trojan asteroids, follow the same orbit as Jupiter, while other Trojans are in the orbits of Neptune, Uranus, Mars, and Earth. (b) Actual positions of all known asteroids at Jupiter's orbit or closer. The locations of the

belt asteroids are indicated by green dots. Objects passing closer than 1.3 AU to the Sun are shown by red circles. Objects observed at least twice are indicated by filled circles, and objects seen only once are indicated by outline circles. Jupiter's Trojan asteroids are deep blue squares. Comets are filled and unfilled light-blue squares. Although the asteroids appear packed together in this drawing, they are typically millions of kilometers apart. The small scale here is deceiving! (b: Minor Planet Center)

between the orbits of Mars and Jupiter. Pallas is even dimmer and smaller than Ceres and has a diameter of only 600 km (375 mi).

Only two more of these asteroids—Juno and Vesta—were found until the mid-1800s, when telescopes improved. Astronomers then began to discover many more asteroids orbiting the Sun at distances from 2 to $3\frac{1}{2}$ AU, between the orbits of Mars and Jupiter. This region of the solar system is now called the **asteroid belt** (Figure 8-8). Asteroids whose orbits lie entirely within this region are called **belt asteroids**.

The spacecraft *Dawn* was sent in 2007 to explore the two most massive belt asteroids Ceres and Vesta, neither of which can be well resolved from Earth. Exquisitely detailed images of Vesta, taken in 2012, reveal that its heavily cratered northern hemisphere is more than a billion years older than the smoother southern hemisphere (Figure 8-9a). Nevertheless, a mountain 20 km (12.4 mi) high pokes up from Vesta's south pole. This mountain is taller than Earth's tallest mountain, Mauna Kea, in Hawaii. When measured from the ocean bottom, Mauna Kea is 4.2 km (2.6 mi) high. Analysis of the asteroid from the data provided by *Dawn* reveals that this as-

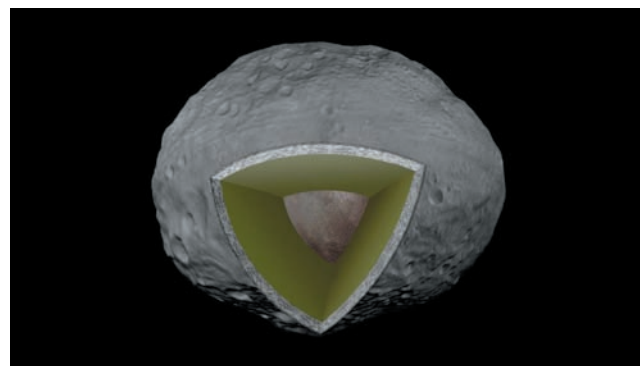
teroid differentiated, meaning that the heavier elements, predominantly iron, settled to the core, while the lighter rock floated upward (Figure 8-9b). *Dawn* will arrive at Ceres in 2015.

Asteroids were discovered one by one throughout the nineteenth century. The next real breakthrough came in 1891, when the German astronomer Max Wolf applied photographic techniques to asteroid searching. A total of 300 asteroids had been found up to that time, each painstakingly discovered by scrutinizing the skies for faint, uncharted "stars" whose positions shifted slowly from one night to the next. With the advent of astrophotography, however, the floodgates were opened. Astronomers could simply aim a camera-equipped telescope at the stars and take long exposures. If an asteroid happened to be in the field of view, it left a distinctive trail on the photographic plate (Figure 8-10). Using this technique, Wolf alone discovered 228 asteroids.

Records of all objects smaller than planets in our solar system are kept by the Minor Planet Center in Cambridge, Massachusetts. The center reports that as of mid-2014, the existence of 644,000 asteroids had been



FIGURE 8-9 Asteroid Vesta (a) Imaged in 2012 by the *Dawn* spacecraft, Vesta's northern and southern hemispheres have different amounts of cratering. The rough area on the bottom is the mountain at the south pole and its environs.



(b) Vesta's internal structure, with iron at the center, surrounded by a mantle of rock (in green) and a thick crust (in gray). (a: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA; b: NASA/JPL-Caltech)

confirmed. Based on these observations, we know that Ceres, the largest asteroid, alone accounts for about 30% of the mass of all known asteroids combined. Only three asteroids—Ceres, Vesta, and Pallas—have diameters greater than 300 km. Indeed, the number of asteroids increases dramatically with decreasing size: Only 41 as-

Insight Into Science

Confirming Observations New findings in science must be confirmed or replicated by other scientists before discoveries or observations are accepted by the scientific community. Therefore, asteroid observations need to be repeated several times to determine exact orbits and eliminate the possibility that a sighting is of a known asteroid or other object.

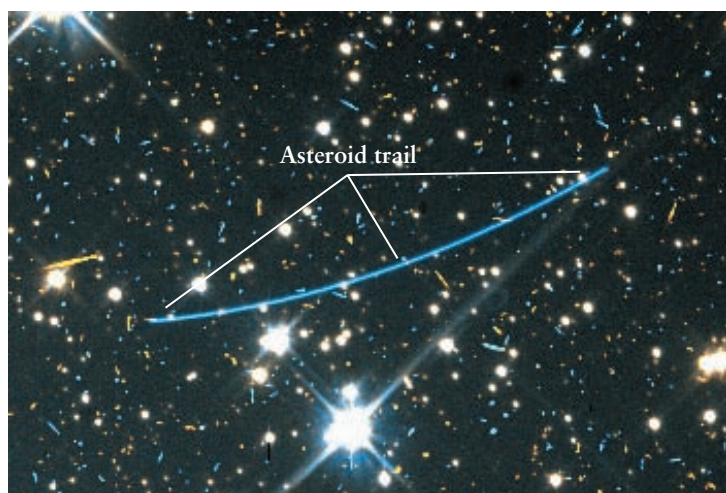


FIGURE 8-10 Discovering Asteroids In 1998, the Hubble Space Telescope found this asteroid while observing objects in the constellation Centaurus. The exposure, tracking stars, shows the asteroid as a 19" streak. This asteroid is about 2 km (1.2 mi) in diameter and was located about 140 million km (87 million mi) from Earth. (R. Evans and K. Stapelfeldt, Jet Propulsion Laboratory and NASA)

teroids have diameters between 200 and 300 km (12 and 190 mi); 250 more are bigger than 100 km (60 mi) across; and there are thought to be tens of millions that are less than 1 km (0.6 mi) across.

1 The Nice model of how the solar system formed explains that most of the debris in the present area of the asteroid belt was ejected by the passage of young Jupiter through this region going both inward and then out again, as discussed in Section 4-3. If all of the present asteroids were collected into a single body, it would have had a diameter of only 1500 km (930 mi), or 12% of Earth's diameter. This is less than two-thirds the diameter of Pluto and half the diameter of our Moon. Like Pluto, such a body would not have enough gravitational attraction to clear the space around it of other bodies. Therefore, a single body that contains the mass of all of the asteroids also would not qualify as a planet.

While Jupiter likely disrupted the mass developing in what is now the asteroid belt some 4.5 billion years ago,

the fact that it has left this region does not mean that its influence there has ended.

8-4 Jupiter's gravity creates gaps in the asteroid belt

In 1867, the American astronomer Daniel Kirkwood called attention to gaps in the asteroid belt. These features, called **Kirkwood gaps**, show the influence of Jupiter's gravitational attraction there today. **Figure 8-11**, a graph of asteroid orbital periods, has gaps at simple fractions ($\frac{1}{3}$, $\frac{2}{5}$, $\frac{3}{7}$, and $\frac{1}{2}$) of Jupiter's orbital period. The gravitational pull of Jupiter created these empty regions by giving objects in them periodic reinforcing tugs that pulled the orbiting bodies to different orbits. The effect is analogous to the gravitational resonance of Mimas creating the Cassini division in Saturn's rings (see Section 7-10).

Focus Question 8-3

Why do astronomers doubt that the asteroid belt was once made up of a single planet?

② It is likely that the asteroid belt contains tens of millions of asteroids, yet the typical separation between “neighboring” asteroids is an impressive 1 million km (600,000 mi). This is quite unlike the image that has

been created by innumerable popular movies and television shows of asteroids so close together that you must dodge them as you fly past.

Despite the large average separation between asteroids, the gravitational influences of Mars and Jupiter

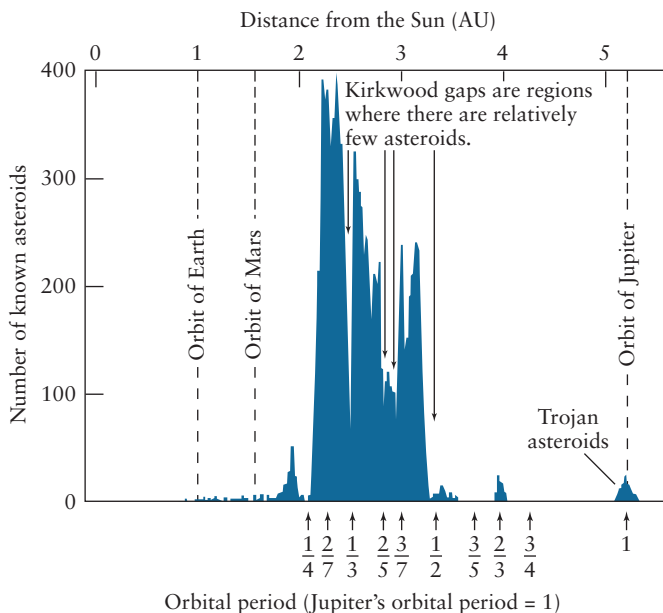


FIGURE 8-11 The Kirkwood Gaps This graph displays the number of asteroids at various distances from the Sun. Note that few asteroids have orbital periods that correspond to such simple fractions as $\frac{1}{3}$, $\frac{2}{5}$, $\frac{3}{7}$, and $\frac{1}{2}$ of Jupiter's orbital period. Resonant orbits with Jupiter have deflected asteroids away from these orbits. The Trojan asteroids accompany Jupiter as it orbits the Sun.

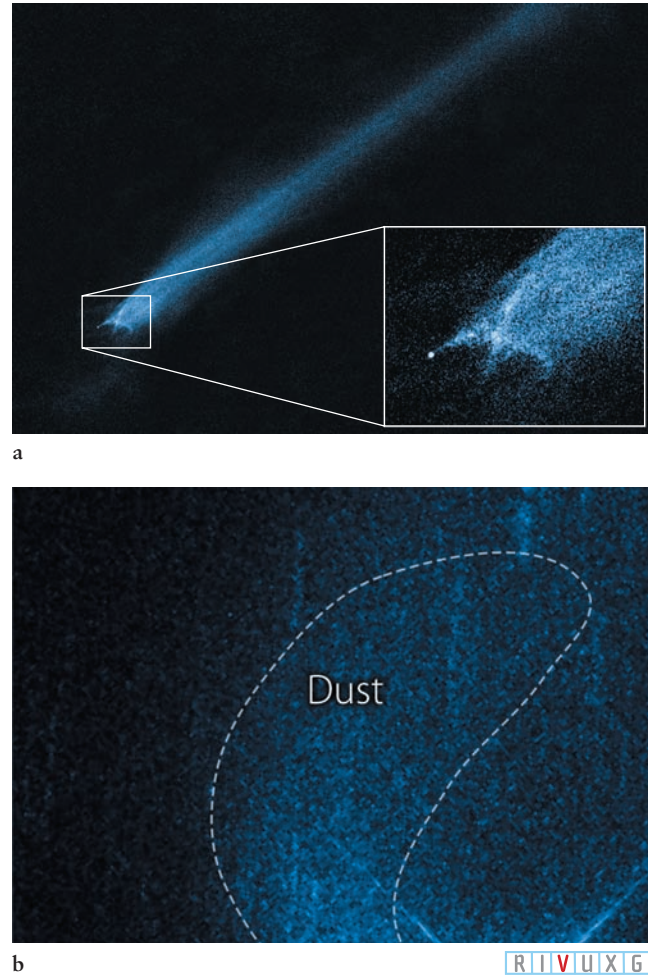
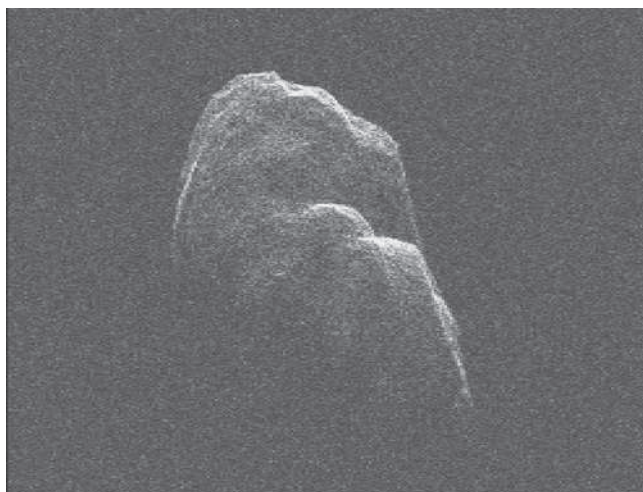


FIGURE 8-12 Collisions Between Two Asteroids (a) Observed in 2010, this X-shaped “object” (inset) is believed to be the collision of an asteroid about 125 m (137 yd) in diameter with a smaller piece of debris, perhaps 10 m (11 yd) across. The event created dust that was pushed away from the Sun, which is to the left and below this image. The collision occurred 2 AU from the Sun and 1 AU from Earth. (b) Seen the same year, asteroid Scheila, about 110 km (68 mi) across, was struck by an asteroid about 30 m (33 yd) across. (NASA, ESA, and D. Jewitt [UCLA])

have sent some asteroids smashing into each other at various times over the past 4.6 billion years. In 2010, astronomers observed two such events (**Figure 8-12**). Typical collision velocities are estimated to be 3600 to 18,000 km/h (2000 to 11,000 mi/h), which is more than sufficient to shatter rock. In some collisions, the resulting fragments may not have enough speed to escape from one another's gravitational attraction, and they reassemble. The asteroid Toutatis (**Figure 8-13**) appears to be composed of two similarly sized pieces connected to each other, as do the asteroids Castalia and Itokawa, among others (see **Figure 8-17**). These bodies are therefore likely to have been broken apart and reformed.



R I V U X G

FIGURE 8-13 Asteroid Toutatis An asteroid whose orbit crosses Earth's, Toutatis was imaged by NASA's Goldstone Solar System Radar in December 2012. (NASA/JPL-Caltech)

Alternatively, several large fragments, such as Ida and Dactyl (**Figure 8-14**), may end up orbiting each other. Dactyl is a pockmarked asteroid some 1.5 km (1 mi) in diameter that orbits Ida at a distance of 100 km (60 mi). Petit-Prince is the satellite of Eugenia. Petit-Prince is 13 km (8 mi) across, orbiting 1200 km (750 mi) from the

Insight Into Science

Backstories Scientists continually apply the “laws of nature” to new situations, problems, observations, experiments, and even popular culture. For example, applying Newton’s law of gravity to the asteroids reveals that they could never swarm, as science fiction movies suggest. At those close quarters, their gravity would cause them either to collide or to pass so close together that they would subsequently fly rapidly and permanently apart.

larger body. Asteroid 87 Sylvia has two known satellites, named Romulus and Remus. To date, 234 asteroid-satellite systems have been observed, including several with two moons.

Some asteroids are just weakly bound agglomerations of rock and metal. In 2013, telescopes on Earth and the Hubble Space Telescope observed that asteroid P/2013 R3 is breaking into pieces (**Figure 8-15**). It is unlikely that a collision caused this activity. A collision would send some debris flying away at great speeds, but this asteroid’s pieces are drifting apart at a leisurely 2 km/h (1 mi/h). Astronomers speculate that the asteroid has probably been weakened by many



R I V U X G

FIGURE 8-14 Ida and Its Satellite The 55-km (34 mi)-long rocky asteroid Ida, shown here with its satellite Dactyl, is about twice the size of the younger asteroid Gaspra (see Figure 4-7). **Inset:** Dactyl is also heavily cratered. (NASA)

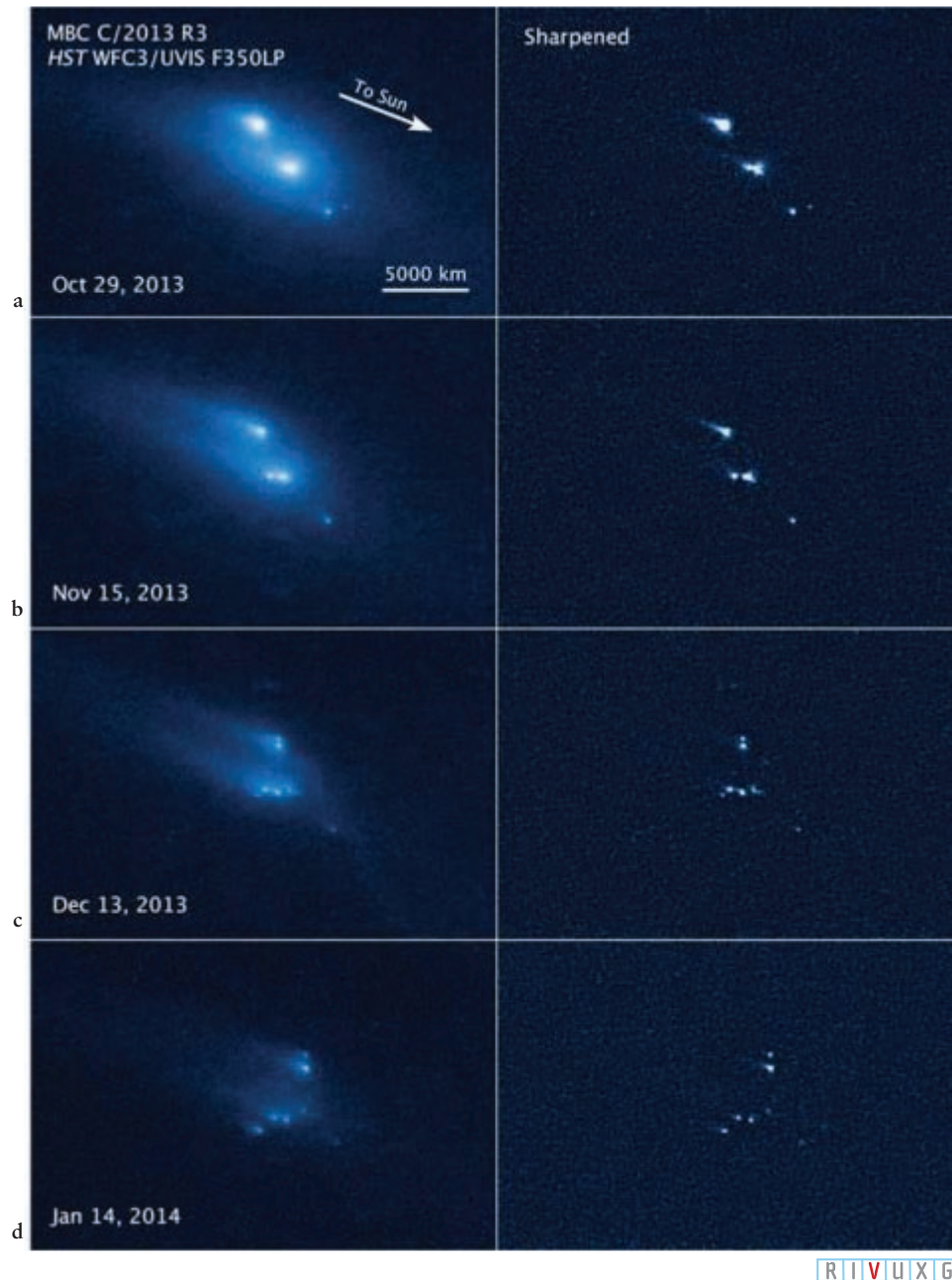


Figure 8-15 Disintegration of an Asteroid Asteroid P/2013 R3 was observed fragmenting into at least 10 pieces late in

2013 and early 2014. This sequence (a)–(d) shows Hubble images of the event. [NASA, ESA, D. Jewitt (UCLA)]

small collisions over billions of years and that heating from the Sun and the asteroid's rotation have finally caused it to fragment.

Another asteroid, P/2013 P5, also discovered in 2013, is rotating and radiating six tails composed of dust from the asteroid body. The tails appear periodically from this 0.43 km (1400 ft) diameter body, which has lost between 100 and 1000 tons of dust, a small fraction of its total mass. Whether this dust is a precursor to the asteroid breaking apart remains to be determined.

In 1918, the Japanese astronomer Kiyotsugu Hirayama drew attention to groups of asteroids that share nearly identical orbits. These fragments are pieces of

parent asteroids. Pursuing this line of research, astronomers in 2002 discovered that two large asteroids collided only a few million years ago (very recently in solar system history) and created about 20 separate families or clusters of smaller asteroids. Asteroids in each of these families orbit together. The largest remnant cluster of this impact is named Karin, after its largest member, an asteroid some 20 km (12 mi) across.

In the early 1990s, the Jupiter-bound *Galileo* spacecraft passed near two asteroids—Gaspia (see Figure 4-7) and Ida (Figure 8-14)—and sent back close-up views. We deduce that Ida is much older because its surface is more heavily cratered than that of Gaspia.

The Sun is not the only star with an asteroid belt. In 2001, astronomers discovered that the star Zeta Leporis, 70 ly from Earth in the constellation Lepus (the Hare), has a disk of debris that appears to contain asteroids. This star system is less than 0.5 billion years old; astronomers hope it will provide insights into the early evolution of the asteroids and other objects in the disk of gas and dust that surrounded the early Sun.

8-5 Asteroids also orbit outside the asteroid belt

While Jupiter's gravitational pull clears out certain areas within the asteroid belt, it actually captures asteroids at two locations in the path of its own orbit. The gravitational forces of the Sun and Jupiter work together to hold asteroids in orbit at these locations, called **stable Lagrange points**, in honor of the French mathematician Joseph Lagrange, whose calculations explained them. One Lagrange point is located 60° ahead of Jupiter, and the other is 60° behind, as shown in **Figure 8-16** (see also Figure 8-8b). Every planet has similar Lagrange points.

The asteroids trapped at Lagrange points are called **Trojan asteroids**. Jupiter's Trojan asteroids are each named after a hero of the Trojan War. As of June 2014, 6073 Trojan asteroids orbiting with Jupiter have been cataloged. Neptune has nine known Trojan asteroids, Mars has five, and Earth and Uranus each have one.

Small solar system bodies orbiting primarily between Jupiter and Neptune are called *centaurs*. The largest known centaur, Chariklo, is especially interesting

because it was observed to have two rings. These were discovered in 2014.

Other asteroids have highly elliptical orbits. Some of these pass through the inner regions of the solar system (see the red dots in Figure 8-8b). The Amor asteroids cross Mars's orbit but do not get as close to the Sun as Earth's orbit, whereas the **Apollo asteroids** do cross Earth's orbit. At least 5766 *Earth-crossing* asteroids are known, several of which are pairs of asteroids that orbit each other. Not all of these Earth-crossing asteroids pose threats to Earth, but at least 1458 of these are *potentially hazardous asteroids* that may someday strike Earth. Still other asteroids have similarly elliptical orbits that extend from the asteroid belt out beyond the farthest reaches of Pluto's orbit.

Several close calls took place in recent years. In 1972, space debris was observed to skip off Earth's atmosphere and retreat back into space. On December 9, 1994, asteroid 1994 XM1 (10 m across—the size of a small bus) passed within 105,000 km (65,000 mi) of Earth. In 2004, asteroid 2004 FU162 (6 m/7 yd in diameter) passed only 6500 km (4000 mi) above Earth's surface. While not as close as this last one, the most spectacular close approach in recent years occurred in 2013, when asteroid 2012 DA14, half the size of a football field, passed 27,700 km (17,200 mi) from Earth (**Figure 8-17**). What made this occurrence especially dramatic was that a large, unrelated piece of space debris that entered the atmosphere on the same day exploded near the Russian town of Chelyabinsk. We discuss that event in more detail shortly. Near misses within the range of these two events happen virtually every year.

During these close encounters, astronomers can examine the details of asteroids. For example, an asteroid's brightness often varies as it rotates because different surface features scatter different amounts of light. Such

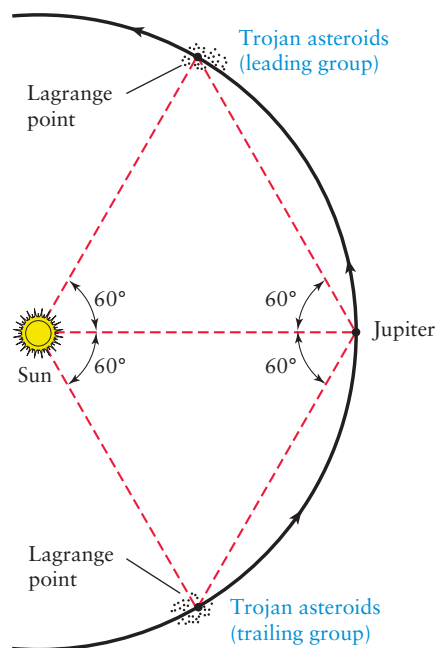


FIGURE 8-16 Jupiter's Trojan Asteroids Groups of asteroids orbit at the two stable Lagrange points along Jupiter's orbit, trapped by the combined gravitational forces of Jupiter and the Sun.

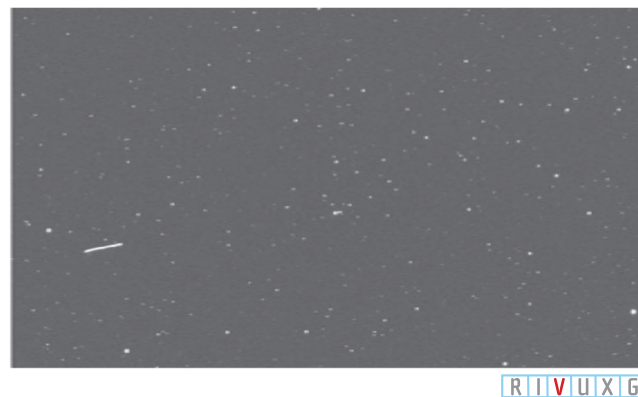


FIGURE 8-17 Asteroid 2012 DA14 This image, taken on February 15, 2013, 7 hours before its closest approach, shows the asteroid as a streak. The camera tracked stars for 3 minutes in order to obtain the image. At closest approach, 2012 DA14 was moving so fast that the same image would have been made in only 15 seconds. (Ernesto Guido/Nick Howes/Remanzacco Observatory)

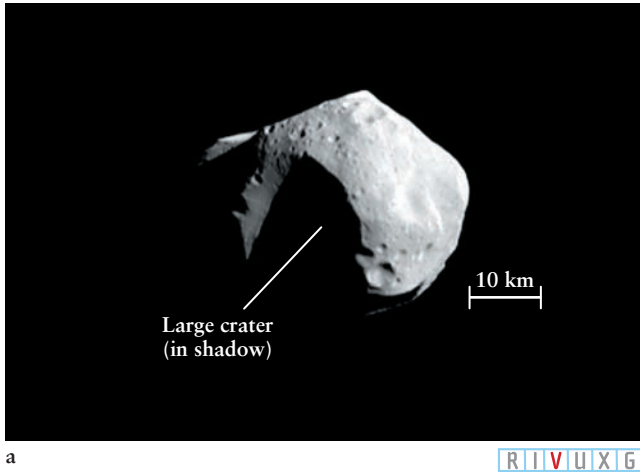
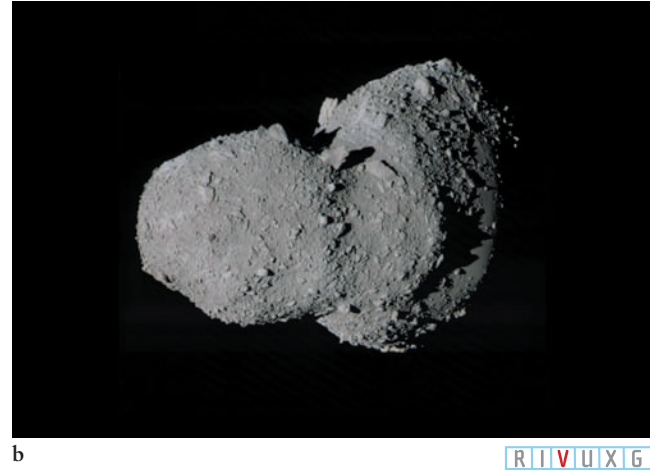


FIGURE 8-18 Asteroids (a) Reflecting only half as much light as a charcoal briquette, Mathilde is half as dense as typical stony asteroids. Slightly larger than Ida (see Figure 8-14), irregularly shaped Mathilde measures $66 \text{ km} \times 48 \text{ km} \times 46 \text{ km}$, ($41 \text{ mi} \times 30 \text{ mi} \times 29 \text{ mi}$) rotates once every 17.4 days, and sitting on Earth, it would weigh 110 trillion tons. The part of the asteroid shown is about $59 \text{ km} \times 47 \text{ km}$ ($37 \text{ mi} \times 29 \text{ mi}$).



The large crater in shadow is about 20 km (12 mi) across. (b) The near-Earth asteroid Itokawa was visited by the Japanese space probe *Hayabusa*. Samples of dust particles from it are now being analyzed. Note the similar two-lobed structure of Itokawa and Toutatis (see Figure 8-13). (a: NASA/JPL/JHUAPL; b: JAXA/Michael Benson/Kinetikon Pictures/Corbis)

data show that typical rotation periods for asteroids are between 5 and 20 hours, although one asteroid, labeled 2010 JL88, rotates about once every 24.5 seconds! This body is the fastest rotating natural object known in the solar system.

The proximity of asteroids and the promise of learning more about these ancient and extremely varied members of our solar system prompted NASA to send the *Near Earth Asteroid Rendezvous (NEAR) Shoemaker*

spacecraft to visit asteroids Mathilde (Figure 8-18a) and Eros. JAXA, the Japanese space agency, sent the *Hayabusa* spacecraft to the asteroid Itokawa (Figure 8-18b), where the spacecraft attempted to take samples. These maneuvers were only partially successful. The *Hayabusa* returned to Earth in June 2010 with its cargo. Dust particles were discovered in the spacecraft. In November 2010, some of the dust particles it carried were confirmed to have come from asteroid Itokawa.

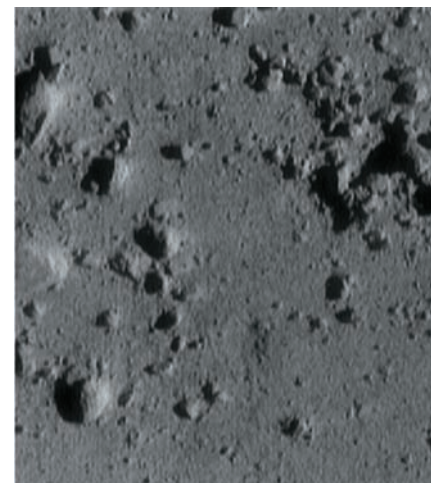
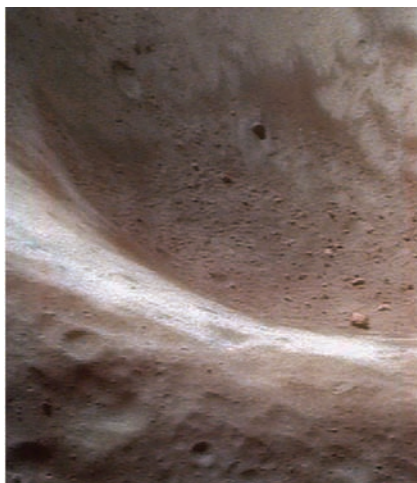


FIGURE 8-19 Asteroid Eros The *Near Earth Asteroid Rendezvous (NEAR) Shoemaker* spacecraft took these images of asteroid Eros in February 1999. (a) The top of the figure is the asteroid's north polar region. Eros's dimensions are $33 \text{ km} \times 13 \text{ km} \times 13 \text{ km}$ ($21 \text{ mi} \times 8 \text{ mi} \times 8 \text{ mi}$) and it rotates every $5\frac{1}{4}$ hours. Its density is 2700 kg/m^3 (4550 lb/yd^3), close to the average density of Earth's crust and twice as dense as

asteroid Mathilde. (b) Looking into the large crater near the top of (a), which is 5.3 km (3.3 mi) across. (c) This is the penultimate image taken by *NEAR Shoemaker* before it gently landed on Eros. Taken from an altitude of 250 m (820 ft), the image is only 12 m (13 yd) across. You can see rocks and boulders buried to different depths in the regolith. (a, b, c: NASA/JPL/JHUAPL)

NEAR Shoemaker revealed that Mathilde is only 1.3 times denser than water and has an albedo of 0.04, making it darker than charcoal. This heavily cratered, carbon-rich body therefore has only half the density of the other asteroids astronomers have studied, such as Ida. Eros is about 3 times as dense as water and rotates once every $5\frac{1}{4}$ hours. In February 2000, *NEAR Shoemaker* went into orbit around Eros, and for a year the spacecraft's cameras and other sensors sent back a wealth of information about it.

NEAR Shoemaker also showed that Eros (Figure 8-19) is a chunk of rock and metal. Whether this asteroid is solid or a collection of loosely held pieces is still under investigation. Analyzing Eros's spectra reveals that it is probably much the same as it was when it coalesced 4.6 billion years ago. Thus, it was never hot enough to differentiate (separate rock from metal). Infrared observations reveal that like our Moon, Eros has regolith. It also has several substantial craters and is strewn with boulders (Figure 8-19c). Astronomers infer from the lack of high crater walls that recent impacts have caused the asteroid to vibrate, collapsing earlier craters. On February 12, 2001, NASA engineers set *NEAR Shoemaker* onto Eros so gently that the spacecraft was able to transmit data after landing. Details of Eros as small as 1.4 m (1.5 yd) across were imaged by the spacecraft.

COMETS

While asteroids consist primarily of rock and metal, other small solar system bodies contain large amounts of frozen water and ices of other compounds. We have seen in earlier chapters that water ice, along with carbon dioxide, methane, and ammonia ices, were locked up in planets and moons. Ices in the young solar system also condensed with roughly equal amounts of small rocky and metallic debris into bodies that still remain in orbit around the Sun. These dirty snowballs in space are the **comets**. The vast majority of them in our solar system are located in the Kuiper belt and Oort cloud (see Figure 4-6).

Comets are named after their discoverers whenever possible. For example, Comet Shoemaker-Levy 9 was the ninth comet discovered by Carolyn and Eugene Shoemaker and separately but simultaneously by David Levy. Today, with large teams of astronomers using sophisticated observing equipment to locate comets, it is sometimes impossible to attribute their discoveries to just a few team members. Comets found under these conditions are named after the team or the research program that finds them.

8-6 Comets come from far out in the solar system

In the first few hundred million years of the solar system's existence, comets formed in its outer reaches at

roughly the distances of Saturn, Uranus, and Neptune. In that region, water was plentiful and the temperature was low enough for the ices to condense into chunks several kilometers across. After they formed, gravitational tugs from Uranus and Neptune flung the comets in every direction, as we saw in Section 4-5.

In 2001, astronomers were able to measure the temperature at which ammonia ice formed in Comet LINEAR (named after the Lincoln Near Earth Asteroid program that discovered it). This temperature determines the ammonia's solid structure, and by determining the structure of the ammonia in Comet LINEAR, scientists were able to calculate the temperature when it formed and therefore how far the comet was from the Sun at that time. That comet coalesced between the present orbits of Saturn and Uranus.

The solar system is believed to contain two reservoirs of comets, the donut-shaped Kuiper belt and the spherical Oort cloud, from which some comets are nudged toward the inner solar system by near misses between them (see Figure 4-6). Indeed, such near misses are the primary source of comets in the realm of the planets.

The Kuiper belt (Figure 8-20) is centered on the plane of the ecliptic. Its main body of objects extends out some 50 AU from the Sun. (Recall that Neptune orbits at 30 AU

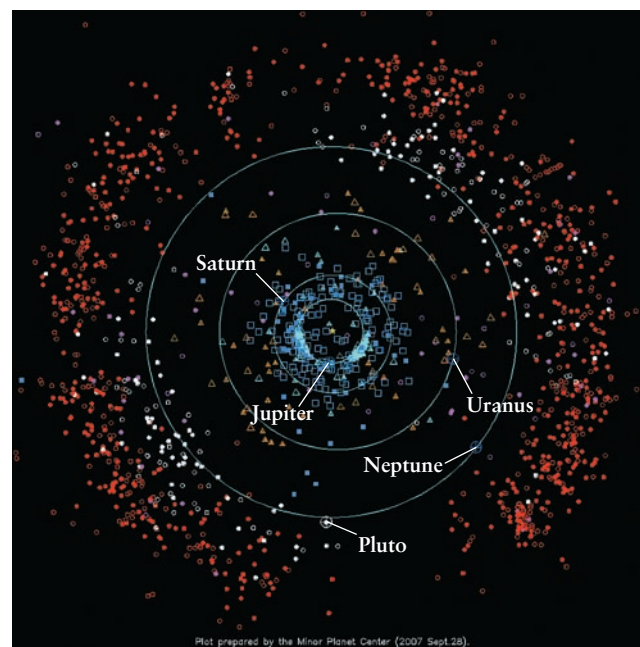


FIGURE 8-20 Current Positions of Known Dwarf Planets and SSBs in the Outer Solar System Objects with unusually high eccentricity orbits are shown as cyan triangles. Objects roaming among the outer planets, called centaur objects, are orange triangles. Plutinos are white circles. Miscellaneous objects are magenta circles, and classical KBOs are red circles. Objects observed only once are denoted by open symbols; objects with two separate observations are denoted by filled symbols. Comets are filled and unfilled light-blue squares. (Courtesy of Gareth Williams, Minor Planet Center)

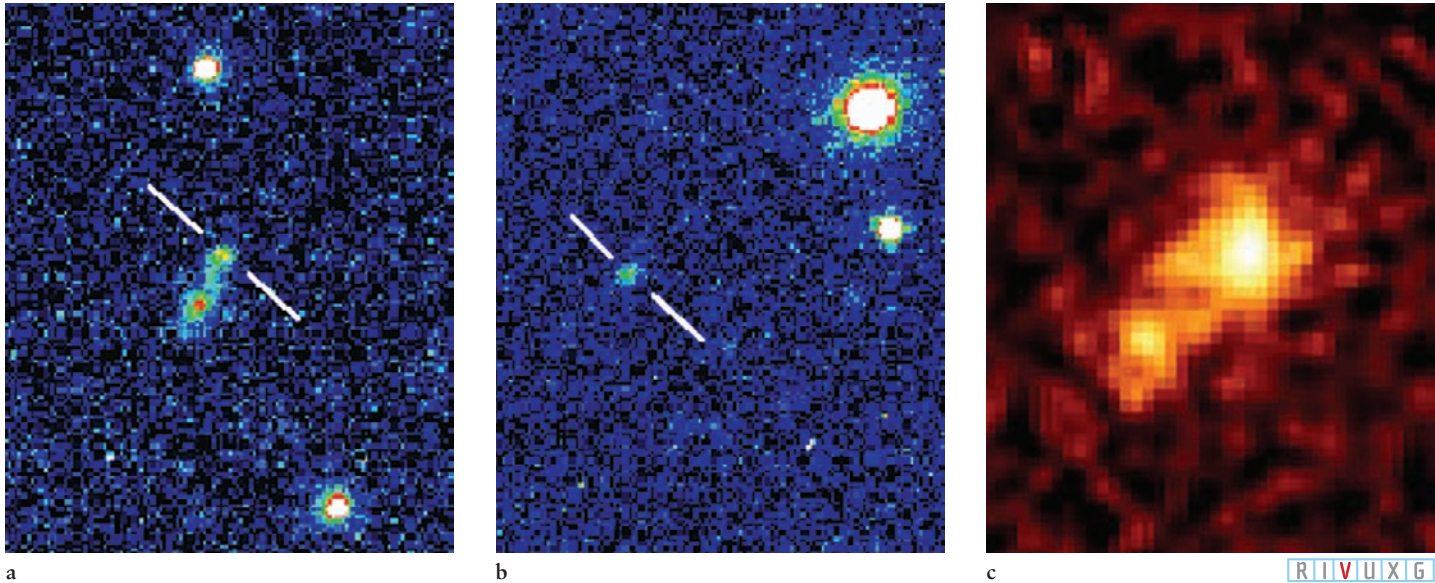


FIGURE 8-21 Kuiper Belt Objects (a, b) These 1993 images show the discovery (bracketed by white lines) of one of at least 1524 known KBOs. These two images of KBO 1993 SC were taken 4.6 hours apart, during which time the object moved

against the background stars. (c) The KBO 1998 WW31 and its moon (lower left). (a, b: Alan Fitzsimmons, Queen's University of Belfast; c: C. Veillet/Canada-France Hawaii Telescope)

from the Sun.) Astronomers believe most comets entering the inner solar system come from the Kuiper belt because most of these comets have orbits relatively close to the plane of the ecliptic. If most comets came from the spherical Oort cloud, the planes of their orbits when near the Sun would be random, with most located far from the plane of the ecliptic.

The Kuiper belt objects were first discovered in 1992 and they fall into two groups. Those with roughly circular orbits are called *classical KBOs* and are found between 30 AU and 50 AU from the Sun. Objects in the second group, called *scattered KBOs*, have more elliptical orbits that range from 35 AU to at least 200 AU from the Sun. More than 1500 KBOs have been observed (Figure 8-20 and Figure 8-21). Although most of the observed KBOs are tens of kilometers across, astronomers have also begun to observe some KBOs only 10 to 100 m in diameter. Based on these latter observations, astronomers estimate that 10^{15} such objects exist out there.

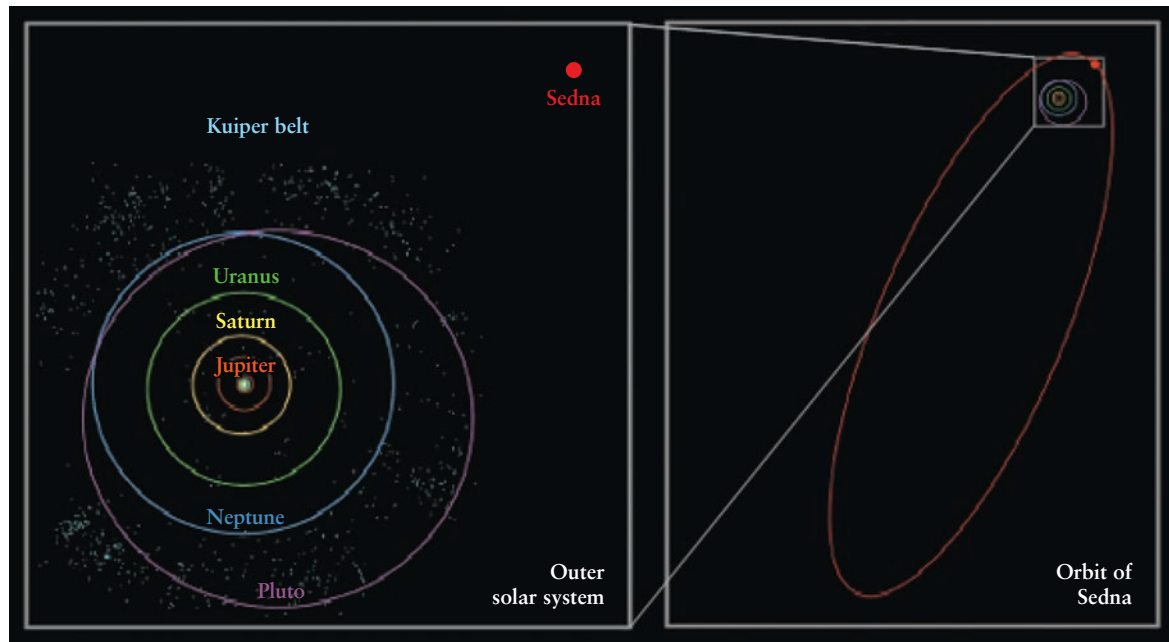
At least 1% of the KBOs are orbiting pairs (Figure 8-21c), including Quaoar, which is 1250 km (775 mi) across, some 400 km (250 mi) larger than Ceres or about half the size of Pluto. At least 122 of the KBOs orbit the Sun in the same region as Pluto. These latter bodies are also called *plutinos* (Section 4-5). The period of Neptune's orbit around the Sun is two-thirds as long as the orbital periods of Pluto and the plutinos. Indeed, it appears that Pluto and the plutinos were forced into their present orbit by Neptune's gravity. Thus, astronomers say that plutinos are locked into a 2:3 *resonance* with Neptune.

The vast majority of comets, estimated between 0.1 and 2×10^{12} bodies, are believed to lie even farther from the Sun. Unlike the Kuiper belt comets and the rest of the solar system, these bodies are believed to have a spherical distribution around the Sun, called the **Oort cloud** (Figure 4-6), named after the Dutch astronomer Jan Oort, who first proposed its existence in the 1950s. The Oort cloud extends from roughly 5000 to 100,000 AU from the Sun. In other words, the outer edge of the Oort cloud is some two-fifths the distance to the star nearest to the solar system.

In 2003, astronomers discovered an object that orbits between the Kuiper belt and Oort cloud. Larger than Quaoar, Sedna is roughly 1800 km (1100 mi) across and is presently 86 AU from the Sun. Sedna has a highly elliptical orbit (Figure 8-22) that takes it from 76 to 937 AU. It has a sidereal rotation period of 10 hours and it is presently the farthest known KBO from the Sun.

Most comets have orbits sufficiently circular that they never even get as close as Pluto is to the Sun. However, occasionally a near miss between two comet nuclei nudges a distant comet toward the inner solar system. As a result of its inward plunge, comets from the Oort cloud, like Hale-Bopp and Hyakutake, have highly elliptical, parabolic, or even slightly hyperbolic orbits (see Section 2-8). Often Oort cloud comets also have orbits that are highly tilted, even perpendicular, to the plane of the ecliptic.

Because the comets in the Kuiper belt and Oort cloud are far from the Sun, they are completely frozen. Solid cometary bodies, called **nuclei** (singular, **nucleus**), are typically between 1 and 10 km (0.6 to 6 mi) across,



a
FIGURE 8-22 Sedna's Orbit (a) The farthest known body in the solar system is in a highly elliptical orbit

b
(b) that orbits between the Kuiper belt and the Oort cloud. (NASA/Caltech)

Focus Question 8-4

What is the solid part of a comet called?

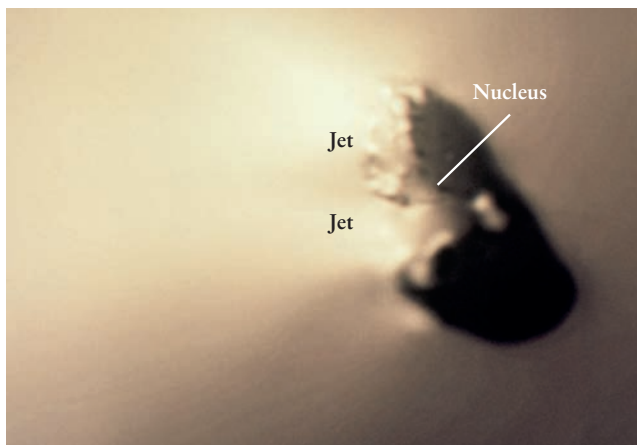
Focus Question 8-5

What is the composition of a comet?

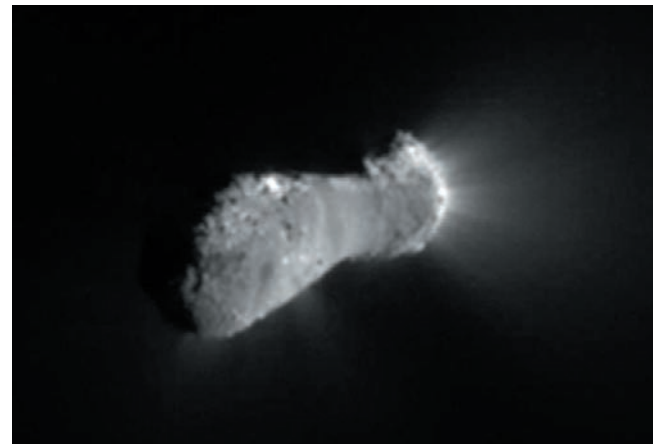
although some, like that of Comet Halley, are 15 km (9 mi) or more in diameter. The first pictures of a comet's nucleus (**Figure 8-23a**) were obtained when a fleet of spacecraft flew past Comet Halley in 1986. Halley's potato-shaped nucleus is darker than coal, probably because of carbon-rich compounds left behind as ice evaporates. In 2010, NASA's EPOXI space-

craft flew past Comet Hartley 2, imaging an overactive body emitting primarily carbon dioxide vapor and water ice out its ends and water vapor out its sides (**Figure 8-23b**). These emissions occur on the sides of the comet facing the Sun. Astronomers believe that the water ice is dragged out by the carbon dioxide gas. All these emissions make Hartley 2 the most complex comet observed so far.

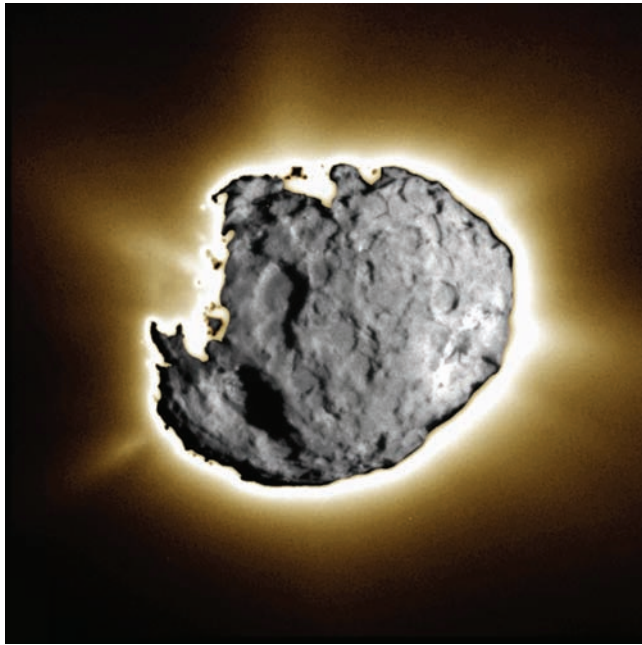
To better determine the composition of the gas emitted by comets, the *Stardust* probe visited Comet Wild 2



a
FIGURE 8-23 Comet Nuclei (a) This image, taken by the *Giotto* spacecraft, shows the potato-shaped nucleus of Comet Halley. Its dark nucleus measures 15 km (9 mi) in its longest dimension and about 8 km (5 mi) in its shortest. The numerous bright areas on the nucleus are icy outcroppings that reflect

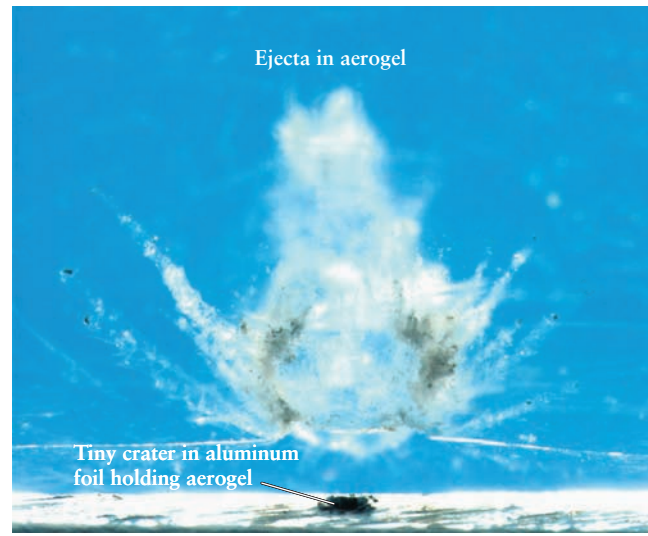


b
more sunlight than surrounding areas of the comet. Two jets of gas can be seen emanating from the left side of the nucleus. (b) The nucleus of Hartley 2 is 2 km (1.2 mi) long. It spins on its short axis once every 18 hours, while tumbling around its long axis. (a: Roger Ressmeyer/CORBIS; b: NASA/JPL-Caltech/UMD)



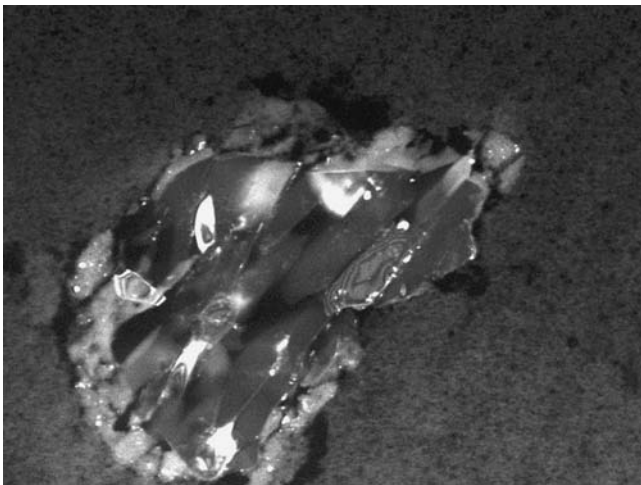
a

R I V U X G



b

R I V U X G



c

R I V U X G

FIGURE 8-24 Comet Wild 2 (a) This picture shows two images combined. One is a high-resolution photograph showing the surprisingly heavily cratered comet. The other image is a longer photograph showing gas and dust jetting from the comet. Its tails are millions of kilometers long. (b) A substance called aerogel was used to capture particles from Comet Wild 2's dust tail. A piece of space debris pierced the aluminum foil holding the aerogel and embedded in it, along with pieces of the foil. (c) A 2- μm piece of comet dust, composed of a mineral called forsterite. On Earth this mineral is used to make gems called peridot. (a, b: NASA/JPL; c: NASA/JPL-Caltech/University of Washington)

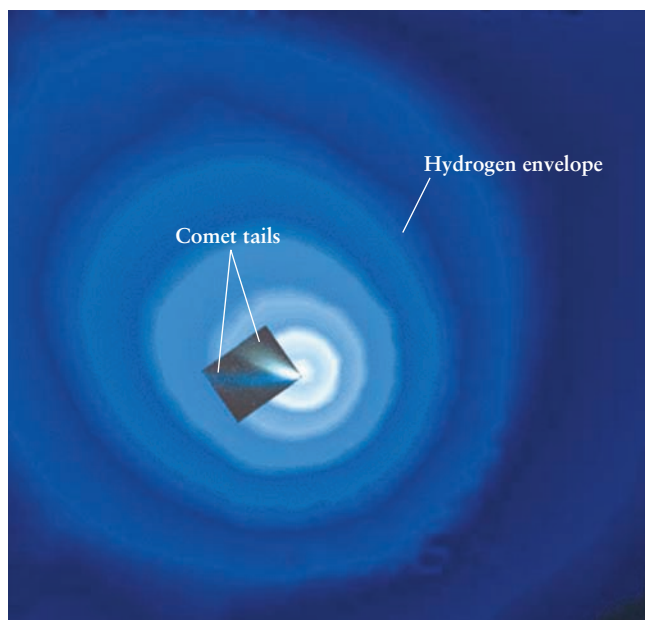
(pronounced *vilt* 2), where it collected samples and returned them to Earth in 2006 (Figure 8-24). Among the most interesting discoveries was the presence of **amino acids**, the building blocks of proteins upon which terrestrial life is based. Comet Wild 2 itself showed many unexpected features, including towering spires, many craters and cliffs, and more than 24 jets, rather than the two or three jets seen on other comets.

When a comet nucleus comes within 20 AU of the Sun, solar radiation begins to sublimate (change directly from solid to gas) the ices on its surface. The liberated gases form an atmosphere, or **coma**, around the nucleus. Because the coma scatters sunlight, it appears as a fuzzy, luminous ball. The largest coma ever measured was that of Comet Holmes, which in

2007 near its perihelion was 1.4 million kilometers (870 thousand mi) across—larger than the Sun. Not visible to the human eye is the **hydrogen envelope**, a sphere of tenuous gas surrounding the comet's nucleus and measuring as much as 20 million km (12 million mi) in diameter (Figure 8-25).

8-7 Comet tails develop from gases and dust pushed outward by the Sun

3 **4** The most visible and inspiring features of comets are their long, flowing, diaphanous tails (Figure 8-26). Comet tails develop from coma gases and dust pushed outward by radiation and particles from the Sun. Thus, comet tails do not trail behind



R I V U X G

FIGURE 8-25 Comet Hale-Bopp In 1997, Comet Hale-Bopp had a hydrogen envelope 1 AU in diameter (blue ovals). This gas was observed in the ultraviolet. The visible-light inset shows the scale of the visible tails (see also the image at the opening of this chapter). (Johns Hopkins University and Naval Research Laboratory)



R I V U X G

FIGURE 8-26 Comet West Astronomer Michael M. West first noticed this comet on a photograph taken with a telescope in 1975. After passing near the Sun, Comet West became one of the brightest comets of the 1970s. This photograph shows the comet in the predawn sky in March 1976. (Dennis di Cicco/Corbis)

the nucleus, as the exhaust from a jet plane does in Earth's atmosphere. Rather, *at the comet's nucleus, the tails always point away from the Sun* (Figure 8-27), regardless of the direction of the comet's motion. The implication that something from the Sun “blows” the comet's gases radially outward led Ludwig Biermann to predict the existence of the solar wind (see Section 6-5 and Section 9-3). This stream of particles from the Sun was actually discovered in 1962, a full decade after it was predicted, by instruments on the spacecraft *Mariner 2*. In 2001, NASA launched the *Genesis* spacecraft, which collected pristine solar wind particles for 2½ years. Although it crash-landed back on Earth in September 2004, some of the particles that it collected were salvaged and studied. From these, astronomers discovered that the chemistry of the Sun and the chemistry of Earth have important differences. In particular, the isotopes (see Section 3-16) of oxygen and nitrogen are found in different amounts on the two worlds. For example, when comparing the amount of ^{14}N (nitrogen with 7 neutrons) and ^{15}N (nitrogen with 8 neutrons), Earth has 40% more ^{15}N than does the Sun. The reasons for this and other chemical differences are now being studied.

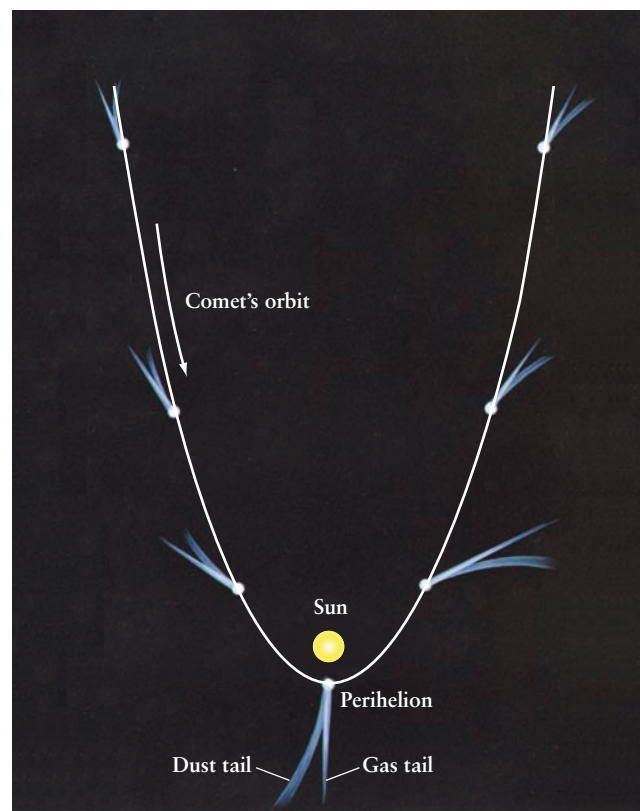


FIGURE 8-27 The Orbit and Tails of a Comet The sunlight and solar wind blow a comet's dust particles and ionized atoms away from the Sun. Consequently, comets' tails always point away from the Sun.

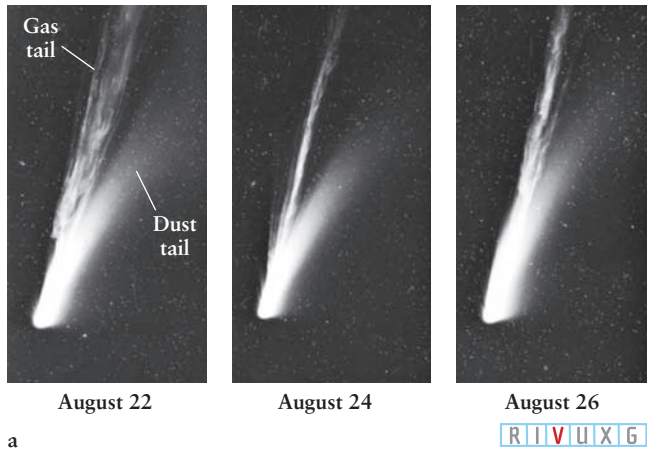


FIGURE 8-28 The Two Tails of Comet Mrkos (a) Comet Mrkos dominated the evening sky in August 1957. These three views, taken at 2-day intervals, show dramatic changes in the comet's gas tail. In contrast, the slightly curved dust tail remained fuzzy and featureless. (b) Wind blowing smoke from



this forest fire causes the smoke column to change shape and direction, just like the solar wind and sunlight cause the tails of comets to change shape and direction. (a: Palomar Observatory; b: Paul Von Baich/The Image Works)

Outflowing particles and radiation from the Sun produce two comet tails: a **gas** (or **ion** or **plasma**) **tail** and a **dust tail** (Figure 8-28 and the figure opening this chapter). Positively charged ions (atoms missing one or more electrons) from the coma are swept directly away from the Sun by the solar wind to form the gas tail. Gas tails are composed primarily of water, carbon monoxide, carbon dioxide, methanol, formaldehyde, and ammonia. Such a tail often appears blue, because carbon monoxide ions in the gas strongly scatter blue light. The relatively straight gas tail can change dramatically from night to night (Figure 8-28a). These changes occur because the solar wind that pushes the gases changes rapidly and the gases themselves are very light and easily moved about, like the smoke from a fire (Figure 8-28b).

The dust tail is formed when sunlight strikes dust particles that have been freed from the comet's evaporating nucleus. Light exerts pressure on any object that absorbs or scatters it. While this pressure, called **radiation pressure** or **photon pressure**, is quite weak, fine-grained dust particles in a comet's coma are sufficiently light to be pushed from the vicinity of the comet, thus

producing a dust tail. The dust tail is often white or light gray. The dust particles are massive enough not to flow straight away from the Sun; rather, the dust tail arches in a path that lies between the gas tail and the direction from which the comet came (see Figure 8-27, Figure 8-28, and Figure 8-29). Even though gas tails of most comets appear significantly brighter and often longer

Insight Into Science

Starting with Observations Science often advances by working from observations and experiments requiring scientific explanations. For example, seeing comet tails and how they behave led Biermann to wonder what causes them. The simplest physical explanation is that matter from each cometary body is being evaporated and pushed upon by something from the Sun—the solar wind. Remember Occam's razor (Section 2-1).

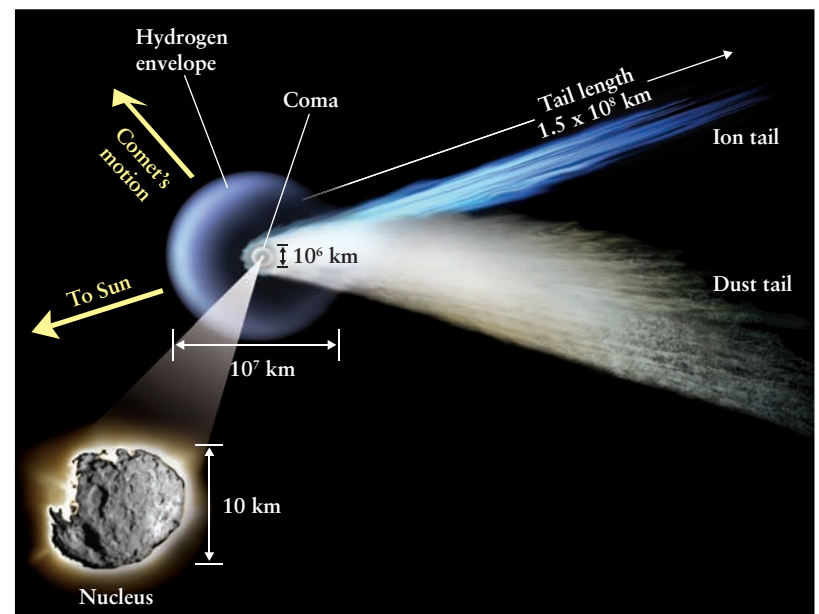


FIGURE 8-29 The Structure of a Comet The solid part of a typical comet (the nucleus) is roughly 10 km in diameter. The coma can be as large as 10^5 to 10^6 km across, and the hydrogen envelope is typically 10^7 km (6 million mi) in diameter. A comet's tail can be enormous—even longer than 1 AU. Comet Wild 2 (inset) is examined further in Figure 8-24. (This drawing is not to scale.) (NASA/JPL)



FIGURE 8-30 The Head of Comet Brooks This comet had an exceptionally large, bright coma. Named after its discoverer, William R. Brooks, it dominated the night skies in October 1911. (Lick Observatory/Science Source)

than dust tails and other rocky debris, observations reveal that about 3 times as much mass leaves a typical comet in the form of dust and rocks than leaves as gas.

Figure 8-29 outlines the elements of comets, although many have bigger or smaller versions of these features.

For example, some comets, like the one shown in **Figure 8-30**, have a large, bright coma but short, stubby tails. Others have an inconspicuous coma but one or more tails of astonishing length. The gas tail in **Figure 8-31** stretches more than 150 million km (1 AU) in length.

Focus Question 8-6

Which comet tail points directly away from the Sun, and why?

8-8 Comets are fragile yet durable

To better understand the chemistry and structure of comets, astronomers sent the *Deep Impact* spacecraft, roughly the size of a small refrigerator, slamming into Comet Tempel 1 at 36,000 km/h (22,000 mi/h) in 2005 (**Figure 8-32a–d**). Observations indicate that the projectile gouged a crater 100 m (110 yd) across and 30 m (33 yd) deep. This collision sent up a cloud of vapor and melted rock estimated to have had a mass of 10 million kg (22 million lb), still less than a ten-millionth of the comet's



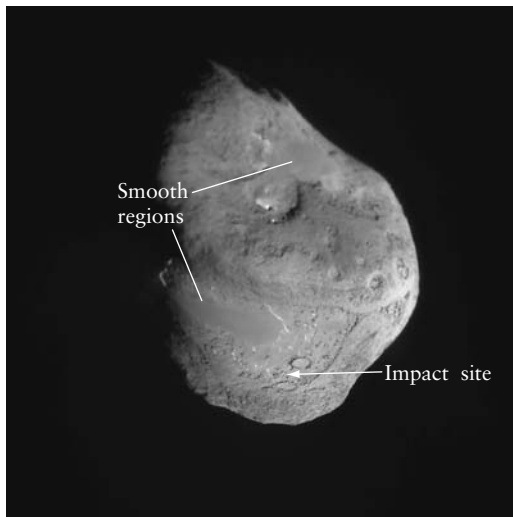
FIGURE 8-31 The Tail of Comet Ikeya-Seki Named after its codiscoverers in Japan, this comet dominated the predawn skies in late October 1965. The yellow in the tail comes from emission by sodium atoms in the dust that was released by the comet. Although its coma was tiny, its tail spanned over 1 AU. (Roger Lynds/NOAO/AURA/NSF)

total mass. The particles ejected were typically smaller than 100 microns (μm).

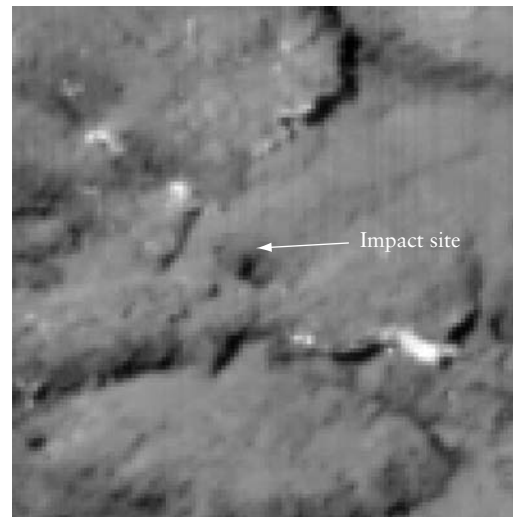
The size of the impact crater and the volume of matter ejected provide strong evidence that this comet is composed of very weakly bound fragments of rock and ice, rather than solid pieces of each. Spectra of the debris thrown upward revealed many organic molecules, that is, molecules containing carbon atoms. This finding supports the conjecture by some astronomers and geologists that comet impacts brought organic molecules to Earth early in our planet's existence.

Even though this comet has an extremely low gravitational force with which to hold itself together, it managed to pull over 90% of the dust that it ejected back onto itself. The craters on this comet and others show that they have been hit by other pieces of space debris throughout their histories. The comet's ability to essentially reform itself after this and other collisions helps explain why such objects have not all been annihilated over time.

To better understand comets and their evolution, the European Space Agency, in cooperation with NASA, sent the *Rosetta* spacecraft to rendezvous with Comet 67P/Churyumov-Gerasimenko (**Figure 8-32e**), which orbits the Sun every 6.45 years. The spacecraft arrived in August 2014. *Rosetta* deployed a lander in November of that year. The data acquired by both the spacecraft and lander are now being studied.



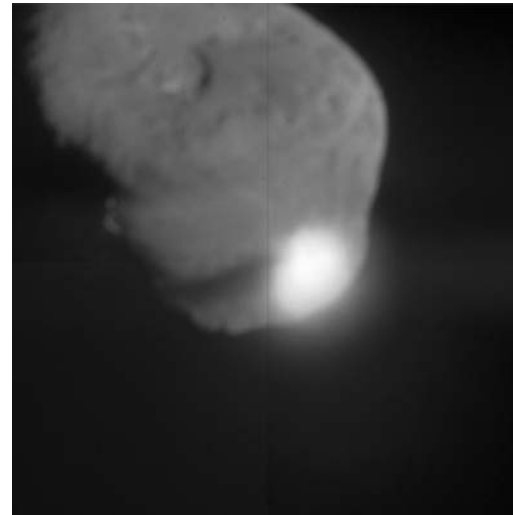
a



b

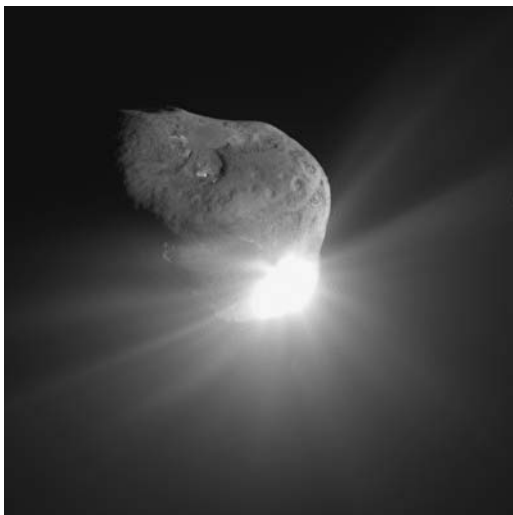


c



d

R I V U X G



e

FIGURE 8-32 Visits to Comet Tempel 1 and Comet 67P (a) This composite image of Comet Tempel 1 has higher resolution at the bottom, as the projectile from *Deep Impact* headed in that direction. The smooth regions on the comet have yet to be explained. (b) Thirty seconds before the projectile struck the comet. (c) Seconds after impact, hot debris explodes away from the comet nucleus. The white horizontal half-ellipses are areas where the CCDs were overloaded with light from the event. (d) An image taken 67 seconds after impact. Within minutes, the cloud of debris became much larger than the entire nucleus. (e) Comet 67P/Churyumov-Gerasimenko as photographed by the *Rosetta* spacecraft in 2014. <http://rosetta.jpl.nasa.gov/gallery/images/comet-67p/churyumov-gerasimenko> (a: NASA/JPL/UMD; b–e: NASA/JPL-Caltech/UMD)

8-9 Comets do not last forever

Astronomers, many of them amateurs, typically discover at least a dozen new comets each year. Falling Sunward from the Kuiper belt, most are **long-period comets**, which have such eccentric orbits that they leave the inner solar system after one pass by the Sun and typically take tens of thousands up to millions of years to return. In 2006, astronomers observed the first comet whose orbit is so elliptical that it is believed to come from the Oort cloud. Called 2006 SQ372, this body has passed the orbit of Neptune on its way out to a maximum distance of 1600 AU from the Sun. This comet will return to the inner solar system in 22,500 years.

Some of these long-period comets can become trapped in the inner solar system. The transformation of their orbits occurs in two different ways. First, a comet can pass so close to a giant planet that the planet's gravitational force profoundly changes the comet's orbit, slowing it down and trapping it in the inner solar system (**Figure 8-33**). Second, the comet can undergo a number of less significant gravitational encounters with planets and other bodies whose cumulative effect is also to trap the comet in the inner solar system. In either case, the long-period comet becomes a **short-period comet**, orbiting the Sun in fewer than 200 years. Like Comet Halley, short-period comets appear again and again at predictable intervals. Halley's Comet has a 76-year orbit and will next pass close to the Sun in 2062.

Comets cannot survive an infinite number of passages near the Sun. A comet's first close encounter causes it to lose a large amount of its surface ice. In doing so, the dust it contains becomes a more significant part of its surface. This dust helps insulate it, reducing the amount of ice and dust the comet loses in subsequent close passages near the Sun. A typical comet is estimated to make between 60 and 100 passages to perihelion (closest approach to the Sun) before losing all of its ice. Because the rock and metal debris in a comet are held together by the ices in which they are imbedded, when the ices are gone, the solid pieces drift apart, many of them remaining in orbit around the Sun. Halley's Comet has been seen at least 27 times, meaning that it has fewer than 73 more close passes to the Sun before it dissipates. Because it returns every 76 years or so, it will survive for about $73 \times 76 = 5548$ more years. A comet can also be torn apart when it comes too close to a planet or the Sun, or it can be destroyed completely by striking a planet, a moon, or the Sun. A spectacular example was Comet Shoemaker-Levy 9. As discussed in Section 7-3, that comet fragmented under the tidal force from Jupiter in 1992. Two years later, with the world's astronomers watching carefully, the pieces returned and struck the planet (see Figure 7-10). Impacts on Jupiter by single objects were observed in 2009 and 2012.

Comet Shoemaker-Levy 9 broke up because its nucleus was held together very weakly, like Comet Tempel

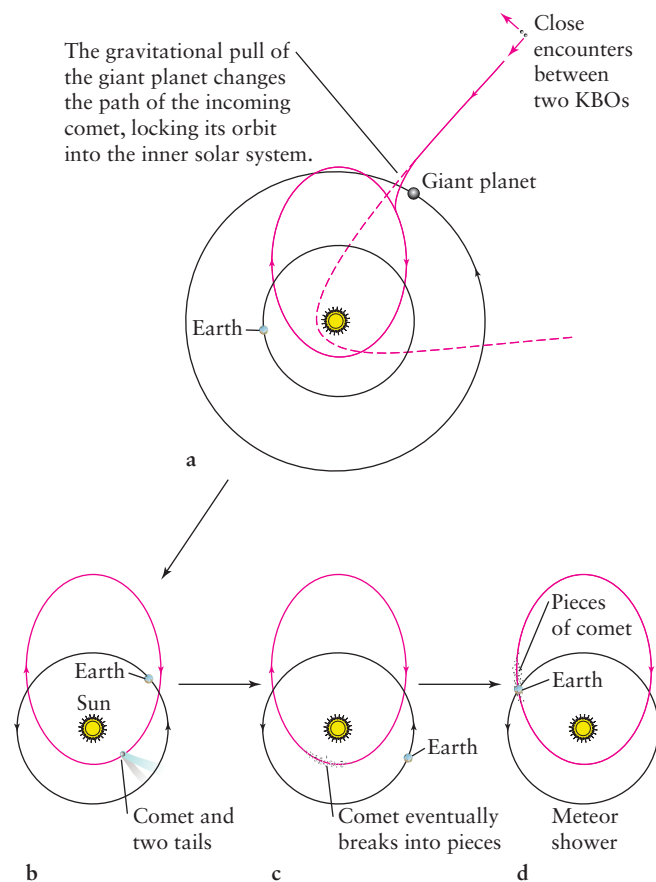
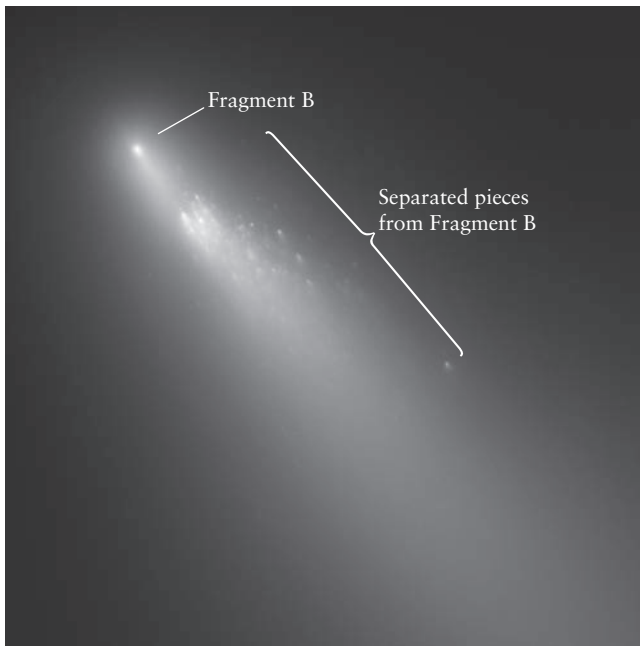


FIGURE 8-33 Transformation and Evolution of a Long-Period Comet (a) The gravitational force of a giant planet can change a comet's orbit. Comets initially on highly elliptical orbits are sometimes deflected into more circular paths that keep them in the inner solar system. (b–d) These figures show the evolution of a comet into gas, dust, and rubble, and why debris from some of these comets strikes Earth (see also Figure 8-40).

1. As discussed previously, Shoemaker-Levy 9 may have been composed of separate pieces that had stuck together until Jupiter's tidal force pulled them apart. Since then, astronomers have observed at least five other comets break up, in 1999, 2000 (Comet LINEAR), 2001, 2002 (Comet Toit-Neujmin-Delpote), and 2006 (Comet Schwassmann-Wachmann 3). **Figure 8-34** shows a piece of Schwassmann-Wachmann 3 coming apart shortly after it passed perihelion. Soon thereafter, its remaining dust and rock fragments spread out in a loose collection of debris that continues to orbit the Sun along the comet's path.

Comets occasionally lose mass quickly by ejecting it in bursts. Several times, starting in September 1995, astronomers observed Comet Hale-Bopp eject 7 to 10 times more mass than usual (**Figure 8-35**). Astronomers think that these events resulted from surface ice and dust being heated by the Sun and then being rapidly ejected from a local region on the comet's surface, called a *vent*. Hale-Bopp rotates with a period of about



R I V U X G

FIGURE 8-34 The Fragmentation of Comet Schwassmann-Wachmann 3 This comet, with a 5.4-year orbit, has been coming apart for decades. In 2006, it further fragmented after passing perihelion. One piece, Fragment B, shed at least 30 smaller pieces, shown here. (NASA; ESA; H. Weaver [APL/JHU]; M. Mutchler and Z. Levay [STLScl])

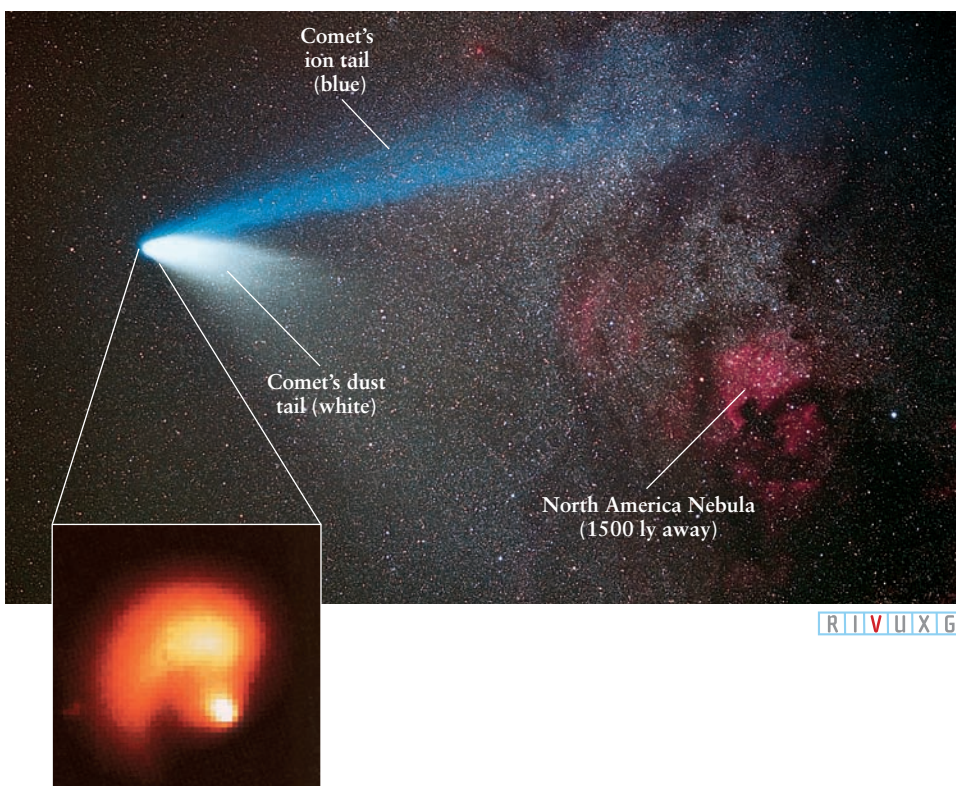
12 hours. The ejected matter therefore formed pinwheel-shaped distributions spiraling from the comet's nucleus at a speed of 109 km/h (68 mi/h). Although the ejected matter did not represent a significant fraction of the comet's mass, its light-reflecting dust made it look like a huge comet fragment.

For thousands of years, astronomers have observed comets that pass very closely to the Sun. This was difficult, of course, because of the Sun's brilliance, so only a few dozen had been carefully observed prior to the mid-1990s. Called *sungrazing comets*, they come within about 0.01 AU (1.46×10^6 km or 9.07×10^5 mi) of the Sun's surface and are typically sublimated and thereby separated into smaller pieces near perihelion. At the end of the nineteenth century, Heinrich Kreutz observed that sungrazers all have the same orbit, indicating that they are pieces of a single comet that broke up perhaps centuries ago. They are called *Kreutz comets* in his honor. In 1995, the Solar and Heliospheric Observatory (SOHO) began a dedicated mission of observing the Sun and, among other things, locating sungrazing comets. To date, it has observed more than 1640 Kreutz comets (Figure 8-36).

The Sun is not the only star that has comets in orbit around it. A star labeled CW Leonis, about 500 ly from Earth, is presently enlarging in size and vaporizing billions of comets around it. Astronomers have observed the spectra of the water vapor from these cometary bodies.

Focus Question 8-7

What evidence do we have that Comet Shoemaker-Levy 9 was not one solid chunk of rock surrounded by ice?



R I V U X G

FIGURE 8-35 Comet Hale-Bopp Discovered on July 23, 1995, this comet was at its breathtaking best in mid-1997. **Inset:** Jets of gas and debris were observed shooting out from Comet Hale-Bopp several times. This image shows the comet nucleus (lower bright region), an ejected piece of the comet's surface (upper bright region), and a spiral tail. The ejected piece eventually disintegrated, following the same spiral pattern as the tail. (Tony and Daphne Hallas, Astrophotos.com)

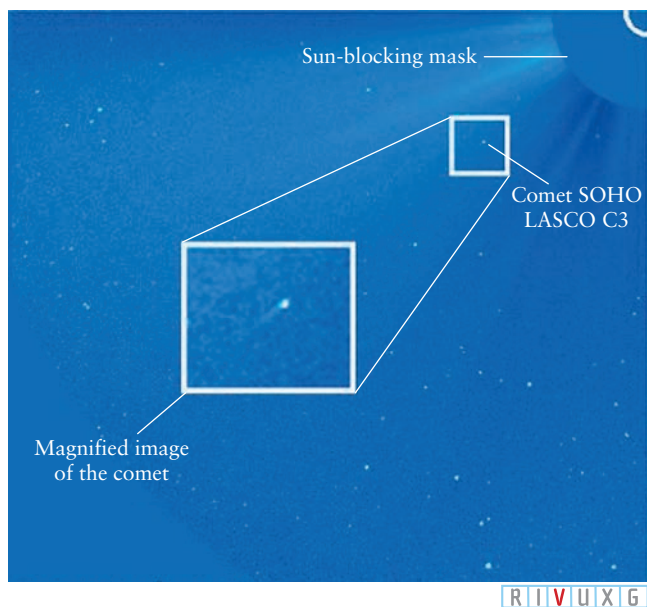


FIGURE 8-36 Sungrazing Comet Comet SOHO LASCO C3 is shown in the smaller box and magnified in the larger one. Discovered in March 2004, it was the 750th sungrazing comet discovered from the SOHO data. It completely sublimated near perihelion. (ESA/NASA-SOHO/LASCO)

METEOROIDS, METEORS, AND METEORITES

As noted earlier, asteroids occasionally collide with each other, sending fragments into interplanetary space.



a



b

These smaller pieces, along with rocky and metallic debris from evaporating comets, and material that never coalesced with larger bodies, are still orbiting the Sun. **Meteoroids** are *rocky and metallic debris smaller than asteroids that are scattered throughout the solar system*. Although no official size standard distinguishes the two, meteoroids are no more than about 10 m (11 yd) across, and the vast majority of them are smaller than 1 mm (0.04 in). **Meteors** are *meteoroids that are being vaporized in the atmosphere*. **Meteorites** are *remnants of meteoroids that land intact*.

By studying the chemistry of various asteroids, astronomers are now beginning to identify which ones were the origins for other asteroids and for meteoroids. For example, the asteroid Braille (named in honor of Louis Braille, developer of the alphabet for the blind) is a piece of debris 2 km (1.2 mi) long that was blasted off the 500-km (310 mi)-diameter asteroid Vesta.

8-10 Small, rocky debris peppers the solar system

5 Passing meteoroids are often pulled by gravity toward Earth. As they move through our atmosphere, they compress and heat the gases in front of them so much that the surrounding air glows. The glowing gas, accompanied by vaporized pieces of the infalling matter, creates the trail we see, and the meteoroid becomes a meteor (**Figure 8-37a**). Common names for these dramatic streaks of light flashing across the sky include *shooting stars*, *fireballs*, and *bolides*. Fireballs are meteors at least



c

FIGURE 8-37 Meteor (a) This brilliant meteor is seen lighting up the dark desert skies of the California desert area of Joshua Tree National Park. Just to the right of the meteor trail are the Pleiades. (b) A spectacular bolide over Chelyabinsk, Russia. The fireball lasted about 30 seconds, leaving a smoke trail about 400 km (250 mi) across the sky, and (c) its impact with the atmosphere produced a powerful shock wave that shattered over a million square feet of glass, and damaged many buildings such as this zinc plant. (a: Atlas Photo Bank/Science Source; b: © Marat Ahmetvaleev/GeoPhoto/ZUMAPRESS.com; c: © Jiang Kehong/Xinhua/ZUMAPRESS.com)

as bright as Venus; bolides are bright meteors that explode in the air. Therefore, “shooting stars” are not stars of any kind or dying stars.

As mentioned in Section 8-5, a spectacular bolide estimated to have been 17–20 m (57–66 ft) across was seen passing Chelyabinsk, Russia, on February 15, 2013 (Figure 8-37b). Moving faster than the speed of sound, the sonic boom it created broke innumerable windows. Flying glass injured about a thousand people. The bolide then exploded at an altitude of about 32 km (20 mi) above nearby Lake Cherbarkul. Had it exploded over the city, the debris that rained down would likely have killed people. As it happened, no one was killed in the event.

One hundred and five years before the meteoritic event in Chelyabinsk, at 7:14 A.M. local time on June 30, 1908, a much more spectacular explosion occurred over Russia, this one over the Tunguska region of Siberia. The blast, comparable to a nuclear detonation of several megatons, knocked a man off his porch some 60 km (37 mi) away and was audible at a distance of more than 1000 km (620 mi). Millions of tons of dust were injected into the atmosphere, darkening the air as far away as California.

Preoccupied with wars, along with political and economic upheaval, neither Russia nor its successor, the Soviet Union, sent a scientific expedition to the site until 1927. At that time, Soviet researchers found that trees had been seared and felled radially outward in an area about 30 km (19 mi) in diameter (Figure 8-38). There was no clear evidence of a crater. In fact, the trees at “ground zero” were left standing upright, although they were completely stripped of branches and leaves. Because no significant meteorite samples were found, for many years scientists assumed that a small comet had exploded in the atmosphere.

Recently, however, several teams of astronomers have argued that a small comet, composed of primarily

light elements and ice, would break up too high above the ground to cause significant damage. They argue that the Tunguska explosion was actually caused by a small asteroid or large meteoroid traveling at supersonic speed. The Tunguska event is consistent with an explosion of an asteroid about 80 m (260 ft) in diameter entering Earth’s atmosphere at a shallow angle, moving at 79,000 km/h (50,000 mi/h) and exploding in the air as a result of becoming exceedingly hot. The resulting shock wave, slamming into the ground, would cause the damage seen without creating a crater. The debate on the cause of the Tunguska event continues.

Most meteors vaporize completely before they can strike Earth. Their dust settles to the ground, often carried by raindrops. (This dust is not the source of acid rain, however, which comes from natural and human-made gases ejected into the atmosphere.) Pieces of a meteor that survive the fiery descent to Earth may leave an **impact crater**. Weather and water erosion are wearing away all of the nearly 200 known impact craters on Earth and thousands more have long since been drawn into Earth by the motion of Earth’s tectonic plates. Indeed, the craters that exist today are all less than 500 million years old because of forces reshaping Earth’s surface.

One of Earth’s least-weathered impact craters is Meteor (or Barringer) Crater near Winslow, Arizona (Figure 8-39). Measuring 1.2 km (0.75 mi) across and 200 m (220 yd) deep, it formed approximately 50,000 years ago when an iron-rich meteoroid some 50 m (55 yd) across (about half the length of a football field) struck the ground at 40,000 km/h (25,000 mi/h). The blast was like the detonation of a 20-megaton hydrogen bomb.

On a typical clear night, you can expect to see a meteor about every 10 minutes. However, at predictable times throughout each year, Earth is inundated with them. These **meteor showers** occur when Earth passes through



FIGURE 8-38 Aftermath of the Tunguska Event In 1908, a stony asteroid traveling at supersonic speed struck Earth’s atmosphere and exploded over the Tunguska region of Siberia. Trees were blown down for many kilometers in all directions from the impact site. (Science Source)



FIGURE 8-39 Meteor Crater An iron meteor measuring 50 m (55 yd) across struck the ground in Arizona 50,000 years ago. The result was this beautifully symmetric impact crater. (D. J. Roddy and K. Zeller/USGS)

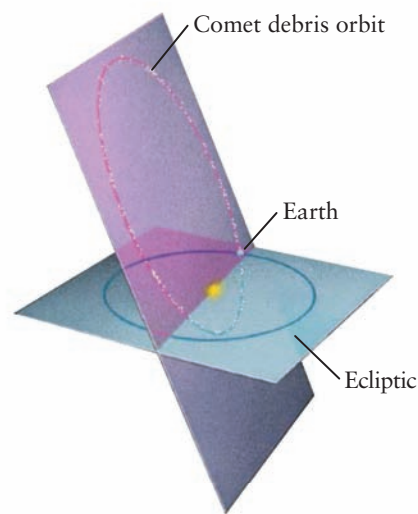


FIGURE 8-40 The Origin of Meteor Showers As comets dissipate, they leave debris behind that spreads out along their orbits. When Earth plows through such material, many meteors can be seen emanating from the same place within a very short time—a meteor shower. As shown in this diagram, many comets have high orbital inclinations.

Focus Question 8-8
Why are the times when meteor showers occur so predictable?

the orbit of debris left behind by a comet (Figure 8-40 and Figure 8-41; also see Figure 8-34 and the photograph on the cover of this book).

Some 30 meteor showers can be seen each year (see Table 8-1). Because the meteors in each shower appear to come from a fixed region of the sky, they are named after the



FIGURE 8-41 Meteor Shower This time exposure, taken in 1998, shows meteors streaking away from the constellation Leo Major. They are part of the Leonid meteor shower. This shower occurs because Earth is moving through debris left by Comet Temple-Tuttle. (Jerry Lodriguss/Science Source)

TABLE 8-1 PROMINENT YEARLY METEOR SHOWERS

Shower	Date of maximum intensity	Typical hourly rate	Constellation
Quadrantids	January 3	40	Boötes
Lyrids	April 22	15	Lyra
Eta Aquarids	May 4	20	Aquarius
Delta Aquarids	July 30	20	Aquarius
Perseids	August 12	80	Perseus
Orionids	October 21	20	Orion
Taurids	November 4	15	Taurus
Leonids	November 16	15	Leo Major
Geminids	December 13	50	Gemini
Ursids	December 22	15	Ursa Minor

constellation from which the meteors appear to radiate (Figure 8-41). For example, meteors in the Leonid shower appear to originate in the constellation Leo Major. More than one meteor can be seen each minute at the peaks of such prodigious meteor showers as the Perseids. These events are among the most exciting light shows in astronomy. Except for the Lyrids, meteor showers are best seen after midnight.

The Moon also moves through the debris that creates meteor showers on Earth. Whereas our atmosphere vaporizes most of the meteors before they strike Earth, the Moon’s thin atmosphere provides no protection, so the Moon is struck by numerous meteoroids during each shower. The Leonid meteor shower of 1999 provided astronomers with an excellent opportunity to observe contemporary impacts on the Moon (see Figure 8-42). Some of these impacts were even bright enough to see with the naked eye. When a 10-kg (22-lb) meteorite strikes the Moon, the impact is as powerful as 10^4 pounds of TNT, and the resulting cloud is momentarily between 5×10^4 and 10×10^4 K, which is hotter than the surface of the Sun. Other meteor showers have provided further impact sightings on the Moon.

8-11 Meteorites are space debris that land intact

Although most meteors that enter Earth’s atmosphere completely vaporize or are destroyed on impact, some land without totally disintegrating. Such pieces of debris are called meteorites, and they come from a variety of sources. By studying their chemistries and comparing them to the chemistries of planets, moons, asteroids,



FIGURE 8-42 Recent Impacts on the Moon The locations A–F are places on the Moon where impacts were observed from Earth in 1999 during the Leonid meteor shower. The impacting bodies hit the Moon at around 260,000 km/h (160,000 mi/h) and had masses of between 1 and 10 kg (2.2 and 22 lb). Each impact created a short-lived cloud that momentarily heated to between 5×10^4 and 10×10^4 K, much hotter than the surface of the Sun. (NASA)

and comets, we know that many meteorites were ejected from these bodies by collisions (Figure 8-43). We saw in Section 6-22, for example, how SNC meteorites offer clues to the chemistry of Mars. By studying the chemistry and age of one meteorite from the Moon that was found on Earth, geologists have determined that it was ejected into space when the Mare Imbrium impact basin formed. Meteorites from the asteroid Vesta show signs of having been molten rock on that world. They were ejected when an impact occurred there, creating a crater 460 km (280 mi) wide and 13 km (8 mi) deep, which has been seen by the Hubble Space Telescope. The spectra also reveal that some meteorites are debris that was never part of larger bodies.

Meteorites tell us the age of the solar system. Measuring the relative amounts of various radioactive elements in them allows astronomers to determine how long ago they formed (see Appendix Q: Radioactivity and the Ages of Objects). The oldest known meteorites became solid bodies about 4.57 billion years ago, and impacts on Earth as early as 4 billion years ago have been identified, based on how they disturbed preexisting rock here. Therefore, the solar system is at least 4.57 billion



FIGURE 8-43 Pieces of the Moon and Mars on Earth (a) Meteorite from the Moon composed of compacted regolith (see Section 6-6). (b) A piece of the Tissint meteorite that landed in Morocco in 2011. Ejected from Mars as a result of an impact 700,000 years ago, it is Martian lava that was weathered on Mars, primarily by water flow. It contains trapped gases from Mars's atmosphere. (a: K. Righter/NASA; b: © Jim Stroepe)

years old, and we will use 4.6 billion years as its age throughout the book.

People have been picking up debris deposited here from space for thousands of years. The descriptions of meteorites in historical Chinese, Indian, Islamic, Greek, and Roman literature show that early peoples placed special significance on these “rocks from heaven.” Meteorites also have a practical “impact”: Infalling space debris is increasing Earth's mass by nearly 300 tons per day on average (Figure 8-44).



FIGURE 8-44 The Mass of Daily Impacts on Earth The Vatican Obelisk is about 300 tons, the amount of mass that strikes Earth daily. As a result, Earth's mass increases by this amount every day. (© Freelance Consulting Services Pty Ltd/Corbis)



a

FIGURE 8-45 Stony Meteorites (a) Most meteorites that fall to Earth are stones. Many freshly discovered specimens, like the one shown here, are coated with thin, dark crusts. This stony meteorite fell in Morocco. (b) Some stony meteorites contain tiny specks of iron, which can be seen when the stones are

Stony meteorites often look like ordinary rocks Meteorites are classified as stones (or stony), stony-irons, and irons. Most **stony meteorites** look much like ordinary rocks, although some are covered with a dark *fusion crust* (Figure 8-45a), created when the meteorite's outer layer melts during its fiery descent through the atmosphere (see Figure 8-41). When a stony meteorite is cut in two and polished, tiny flecks of iron are sometimes found in the rock (Figure 8-45b).

Iron meteorites are very dense and look noticeably different from rocks Iron meteorites (Figure 8-46) may also contain from 10% to 20% nickel by weight. Iron is moderately abundant in the universe as well as being one of the most common rock-forming elements, so it is not surprising that iron is an important constituent of asteroids and meteoroids. Another element, iridium, is common in the iron-rich minerals of meteorites but rare in ordinary rocks on Earth's surface. Iridium binds strongly to iron, so when Earth was first forming and iron was descending into our planet's core, it carried down most of the iridium that was associated with it. Measurements of iridium in Earth's crust can thus tell us the rate at which meteoritic material has been deposited over the ages.

Meteorites with a high iron content can be located with a metal detector. They also look unusual and hence are more likely to be noticed. Consequently, the easily found iron and stony-iron meteorites dominate most museum collections. In 1808, Count Alois von Widmanstätten, director of the Imperial Porcelain Works in



b

cut and polished. This specimen was discovered in Australia. (a: C. Bevilacqua/De Agostini/Getty; b: © Photo by Daniel Ball/School of Earth and Space Exploration/ASU. Courtesy of the ASU Center for Meteorite Studies)



a

RIVUXG

FIGURE 8-46 Iron Meteorites (a) Irons are composed almost entirely of iron-nickel minerals. The surface of a typical iron is covered with depressions resembling thumbprints created as the meteorite's outer layers vaporized during its high-speed descent through the atmosphere. This specimen was found in Argentina.



b

RIVUXG

(b) When cut, polished, and etched with a weak acid solution, most iron meteorites exhibit interlocking crystals in designs called Widmanstätten patterns. This meteorite was found in Australia. (a: RGB Ventures LLC dba SuperStock/Alamy; b: NASA)

Vienna, discovered a conclusive test for the most common type of iron meteorite. Most iron meteorites have a unique structure of long nickel-iron crystals, called **Widmanstätten patterns**, which become visible when the meteorites are cut, polished, and briefly dipped into a dilute solution of acid (Figure 8-46b). Because nickel-iron crystals can grow to lengths of several centimeters only if the molten metal cools slowly over many millions of years, Widmanstätten patterns are never found in counterfeited meteorites (or “meteorwrongs”).

Stony-iron meteorites are the most exotic of all space debris on Earth The final category of meteorites are the stony-iron meteorites, which consist of roughly equal amounts of rock and iron. **Figure 8-47**, for example, shows the mineral olivine suspended in a matrix of iron.



RIVUXG

FIGURE 8-47 Stony-Iron Meteorite Stony-irons account for about 1% of all meteorites that fall to Earth. This specimen, a variety called a pallasite, was found in Antarctica. It has been sliced and polished. (Walter Geiersperger/Corbis)

To understand why different types of meteorites exist, we consider their formation. Many meteorites were once pieces of asteroids. Heat from impacts and from the rapid decay of radioactive isotopes melted newly formed asteroid interiors. Over a few million years after they formed, differentiation occurred, just as in young Earth. Iron sank toward the asteroid's center, while lighter rock floated up to its surface. Iron meteorites are fragments of asteroid cores, and stony-irons come from the boundary regions between the iron cores and stony crusts.

Stony meteorites have a variety of origins. Some come from the outer layers of asteroids. Other stony meteorites, ordinary **chondrites**, show no evidence of ever having been melted as parts of asteroids. They may therefore be primordial material from which our solar system was created. **Carbonaceous chondrites** are rare chondrites that contain small glass-rich beads called *chondrules*. These meteorites also contain complex carbon compounds, including simple sugars and glycerin. They also have as much as 20% water bound into their minerals. The organic compounds would have been broken down and the water driven out if these meteorites had been significantly heated as during large asteroid formation. The small, undifferentiated asteroid Mathilde, shown in Figure 8-18a, has a very dark gray color and virtually the same spectrum as a carbonaceous chondrite meteorite, so it is also likely composed of this primordial material.

Amino acids are among the organic compounds occasionally found inside carbonaceous chondrites. By observing their spectra, we know that organic compounds exist on asteroids and smaller space debris. Indeed, some

Focus Question 8-9

Why are stony-iron meteorites so rare compared to stony or iron meteorites?

PERCENTAGES OF METEORITES THAT LAND AND THAT ARE FOUND			
TABLE 8-2			
	Stony	Iron	Stony-Iron
Meteorites that land	94%	4.8%	1.2%
Meteorites that are found	53%	42%	5%

scientists suspect that amino acid-rich carbonaceous chondrites may have played a role in the origin of life on Earth.

Stony meteorites account for about 94% of all meteoritic material that falls on Earth. Most stony meteorites are overlooked because they look like ordinary stones. Indeed, the percentages of the stony, iron, and stony-iron meteorites that are discovered are quite different from the percentages that actually land. Nevertheless, astronomers and geologists have a good idea about how many of each type strike land. They obtain the correct percentages of impacts by carefully surveying areas in which only meteorites land, namely, snow and ice-covered regions, such as Antarctica, or on deserts. By counting all of the debris under the surface using metal detectors and other technologies, the actual number of impacts of each type of meteorite is determined. Table 8-2 lists the relationship between the number of meteorites that land and those that are found.

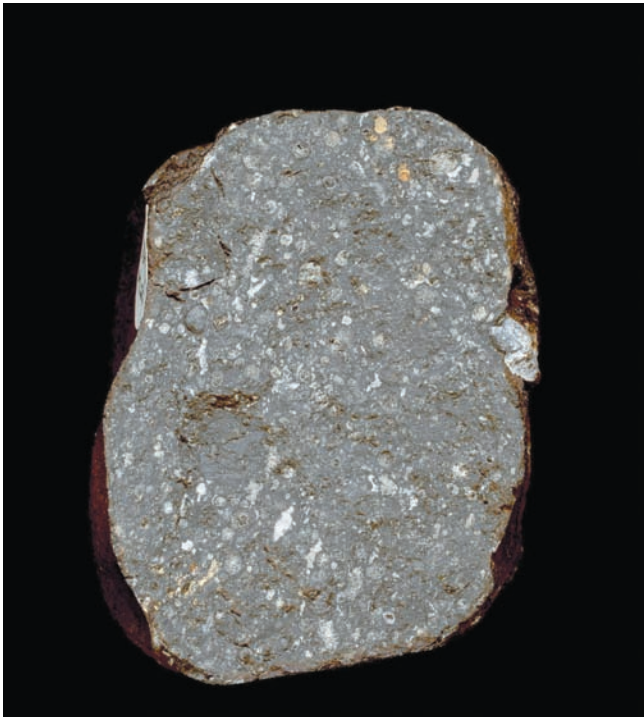
8-12 The Allende meteorite provides evidence of catastrophic explosions

A chance to study debris immediately after a meteorite impact came shortly after midnight on February 8, 1969, when a brilliant, blue-white light shot across the night sky over Chihuahua, Mexico. Hundreds of people witnessed the dazzling display. The light disappeared in a spectacular and deafening explosion that dropped thousands of rocks and pebbles over the terrified onlookers. Within hours, teams of scientists were on their way to collect specimens of carbonaceous chondrites, collectively named the *Allende meteorite*, after the locality in which they fell (Figure 8-48).

One of the most significant discoveries to come from the Allende meteorite was evidence of the detonation of a nearby supernova 4.6 billion years ago (see Section 4-2). Among nature’s most violent and spectacular phenomena, a *supernova explosion* occurs when a massive star blows apart in a cataclysm that hurls matter outward at tremendous speeds, as we will explore in Chapter 12. During this detonation, violent collisions between atomic nuclei produce a host of radioactive elements, including a short-lived radioactive isotope of aluminum. Based on its decay products, scientists found



a



b

RIVUX6

FIGURE 8-48 Pieces of the Allende Meteorite (a) This carbonaceous chondrite fell near Chihuahua, Mexico, in February 1969. Note the meteorite’s dark color, caused by a high abundance of carbon. Geologists believe that this meteorite is a specimen of primitive planetary material. The ruler is 15 cm (6 in) long. (b) Sliced open, the Allende meteorite shows round, rocky inclusions called chondrules in a matrix of dark rock. (a: J. A. Wood; b: The Natural History Museum/The Image Works)

unmistakable evidence that this isotope once lay within the Allende meteorite. Some astronomers interpret this as evidence for a supernova in our vicinity at about the time the Sun was born. By compressing interstellar gas and dust, the supernova’s shock wave may have helped stimulate the birth of our solar system.

8-13 Asteroid impacts with Earth have caused mass extinctions

In the late 1970s, the geologist Walter Alvarez and his father, physicist Luis Alvarez, discovered the remnants of a truly Earth-shaking impact. Working at a site of exposed marine limestone in the Apennine Mountains in Italy that had been on Earth's surface 65 million years ago, the Alvarez team discovered an exceptionally high abundance of iridium in a dark-colored layer of clay between limestone strata (Figure 8-49). Recall that the iridium that was present on the surface when Earth differentiated sank to the core. The iridium found by the Alvarez team had to have been deposited much later.

Since 1979, when this discovery was announced, a comparable layer of iridium-rich material has been uncovered at numerous sites around the world. In every case, geologic dating reveals that this apparently worldwide iridium-rich layer was deposited about 65 million years ago. Paleontologists were quick to realize the significance of this date, because it was back then that all of the dinosaurs rather suddenly became extinct. In fact, two-thirds of all animal species disappeared within a brief span of time 65 million years ago.

The Alvarez discovery supported an astronomical explanation for the dramatic extinction of so much of the life that once inhabited our planet—an asteroid impact. There is no universal agreement on the disasters that befell Earth during this episode. One scenario has an asteroid 10 km (6.2 mi) in diameter slamming into Earth at high speed and throwing enough dust into the atmosphere to block out sunlight for several years. As the temperature dropped drastically and plants died for lack of sunshine, the dinosaurs would have perished, along with many other creatures in the food chain that



RIIVUIG

FIGURE 8-49 Iridium-Rich Layer of Clay This photograph of strata in the Apennine Mountains of Italy shows a dark-colored layer of iridium-rich clay sandwiched between white limestone (bottom) from the late Mesozoic era and grayish limestone (top) from the early Cenozoic era. The coin is the size of a U.S. quarter. (W. Alvarez)

were highly dependent on vegetation. The dust eventually settled, depositing an iridium-rich layer over Earth. Another theory posits that the impact created enough heat to cause planet-wide fires, followed by changes in the oceans that killed many species of life in them. Whatever scenario turns out to be correct, small creatures capable of ferreting out seeds and nuts were among the animals that managed to survive this holocaust, setting the stage for the rise of mammals and, consequently, the evolution of humans.

In 1992, a team of geologists discovered that the hypothesized asteroid crashed into a site on the eastern edge of Mexico. They based this conclusion on glassy debris and violently shocked grains of rock ejected from the multiringed, 195-km (120 mi)-diameter Chicxulub (pronounced *Chih-chuh-lube*) Crater buried under the Yucatán Peninsula and western Caribbean Sea in Mexico (Figure 8-50). From the known rate at which radioactive potassium decays, scientists have pinpointed the date when the asteroid struck—64.98 million years ago. In 1998, geologists digging on the Pacific Ocean floor discovered a piece of meteoritic debris with precisely the same age, apparently a piece of the offending asteroid. While some geologists and paleontologists are not yet convinced that an asteroid impact led to the extinction of the dinosaurs, most agree that this hypothesis better fits the available evidence than any other explanation that has been offered so far.

By comparing the chemical composition of the debris at Chicxulub with the chemical composition of meteorites, astronomers and geologists in 2007 determined that the asteroid that caused the mass extinction 65 million years ago came from a group of asteroids called Baptistinas. This group was created as a result of a larger body, some 170 km (105 mi) in diameter, breaking up after a collision 160 million years ago.

Another impact may have led to an earlier, more devastating mass extinction The end of the dinosaurs' reign was not the only mass extinction caused by an impact. In 2001, evidence came to light concerning the Permian-Triassic boundary of 250 million years ago that suggests another devastating blow from space. That time is called the "Great Dying," a mass extinction during which some 80% of the species of life living on land and 90% of those living in the oceans perished. Rocks from that time discovered in places from Japan to Hungary show evidence, in the form of *fullerenes*—soccer ball-shaped molecules containing at least 60 carbon atoms—of an impact from space. Trapped inside these fullerenes were gases that could only have been forced into them by exploding stars. (This signature of gas-filled fullerenes from space has now also been discovered in the layer of rock that existed on Earth's surface 65 million years ago.)

Focus Question 8-10

What evidence do astronomers have from the Allende meteorite that a supernova explosion occurred at about the time the solar system formed?

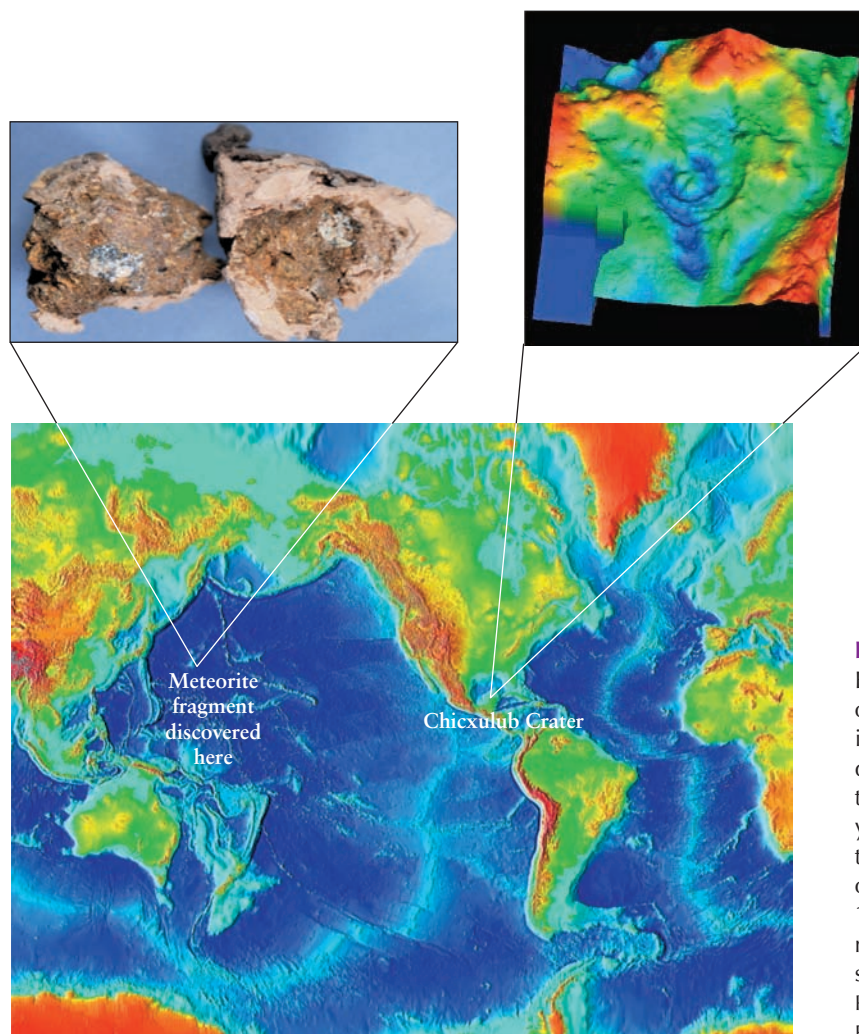


FIGURE 8-50 Confirming an Extinction-Level Impact Site

By measuring slight variations in the gravitational attraction of different materials under Earth's surface, geologists create images of underground features. Concentric rings of the underground Chicxulub Crater (**right inset**) lie under a portion of the Yucatán Peninsula. This crater has been dated to 65 million years ago and is believed to be the site of the impact that led to the extinction of the dinosaurs. A piece of 65-million-year-old meteorite discovered in the middle of the Pacific Ocean in 1998 is believed to be a fragment of that meteorite. The fragment, about 0.3 cm (0.1 in.) long, was cut into two pieces for study (**left inset**). (Peter W. Sloss, NOAA/NGDC; right inset: Mark Pilkington/Geological Survey of Canada/Science Source; left inset: Frank T. Kyte, UCLA)

A 250-million-year-old crater that may have been the relevant impact site was found in 2004 off the northwestern coast of modern-day Australia. Keep in mind that both this impact and the one that ended the reign of the dinosaurs occurred on a world whose surface is in continual motion. Therefore, their present locations are not the same as their initial impact sites. The Chicxulub impact actually occurred in the Pacific Ocean, whereas the Australian impact occurred before the continents were separated from Pangaea (see Section 6-3). The impact explanation of the Great Dying is not universally accepted. Another possible explanation is a massive amount of volcanic eruption that darkened the skies and changed the climate.

Could a catastrophic impact happen again? So many asteroids cross Earth's path that scientists agree that it is a matter of "when" rather than "if." NASA and other agencies are actively cataloging potentially dangerous asteroids. The good news is that studies of craters show that larger asteroids strike Earth significantly less often

than do smaller ones. Although asteroids large enough to create Meteor Crater strike Earth about once every 10,000 years, killer asteroids, like the one that killed off the dinosaurs, collide with Earth only once every 100 million years. The threat of a catastrophic impact by an asteroid or comet in our lifetimes, thankfully, is remote.

SUMMARY OF KEY IDEAS

Asteroids

- Pieces of solar system debris larger than about 10 m (11 yd) and composed primarily of rock and metal are called asteroids.
- Tens of thousands of belt asteroids with diameters larger than a kilometer are known to orbit the Sun between the orbits of Mars and Jupiter. The gravitational attraction of Jupiter depletes certain orbits within the asteroid belt. The resulting Kirkwood gaps occur at simple fractions of Jupiter's orbital period.
- A planet's gravity and the Sun's gravity can combine to capture Trojan asteroids in two locations, called

Focus Question 8-11

How did the Chicxulub Crater move from the Pacific to the Caribbean?

stable Lagrange points, along a planet's orbit. Jupiter, Neptune, Uranus, Mars, and Earth are known to have Trojan asteroids.

- The Apollo asteroids move in highly elliptical orbits that cross the orbit of Earth. Many of these asteroids will eventually strike the inner planets.
- A belt asteroid (Ceres) and four KBOs (Pluto, Eris, Haumea, and Makemake) are classified as dwarf planets.
- Pluto, a KBO and dwarf planet, is an icy world that may well resemble the moon Triton.

Comets

- Comet nuclei are fragments of ice and rock often orbiting at a great inclination to the plane of the ecliptic.
- Many comet nuclei orbit the Sun in the Kuiper belt, a doughnut-shaped region beyond Neptune. Trillions of cometary nuclei are also believed to exist in the spherical Oort cloud located far beyond the Kuiper belt. Some get nudged into orbits taking them close to the Sun, which is when we see their tails.
- In the Kuiper belt and Oort cloud, comets have fairly circular orbits. When close to the Sun, they generally move in highly elliptical orbits.
- As an icy comet nucleus approaches the Sun, it develops a luminous coma surrounded by a vast hydrogen envelope. A gas (or ion) tail and a dust tail extend from the comet, pushed away from the Sun by the solar wind and radiation pressure.

Meteoroids, Meteors, and Meteorites

- Boulder-sized and smaller pieces of rock and metal in space are called meteoroids. When a meteoroid enters Earth's atmosphere, it produces a fiery trail, and it is then called a meteor. If parts of the object survive the fall, the fragments that reach Earth's surface are called meteorites.
- Meteorites are grouped into three major classes according to their composition: iron, stony-iron, and stony meteorites. Rare stony meteorites, called carbonaceous chondrites, may be relatively unmodified material from the primordial solar nebula. These meteorites often contain organic hydrocarbon compounds, including amino acids.
- Fragments of rock from comets that have vaporized all their ices produce meteor showers.
- An analysis of the Allende meteorite suggests that a nearby supernova explosion may have been involved in the formation of the solar system some 4.6 billion years ago.
- An asteroid that struck Earth 65 million years ago probably contributed to the extinction of the dinosaurs and many other species. Another impact may have caused the Great Dying of life 250 million years ago. Such devastating impacts occur on average every 100 million years.

WHAT DID YOU THINK?

- 1 *Are the asteroids a former planet that was somehow destroyed? Why or why not?* No. The gravitational pull from Jupiter prevented a planet from ever forming in the asteroid belt. Also, the present-day total mass of the asteroids is much less than even the mass of tiny Pluto, a dwarf planet.
- 2 *How far apart are the asteroids on average?* The distance between asteroids averages 1 million km (620,000 mi).
- 3 *How are comet tails formed? Of what are they made?* Ices in comet nuclei are turned into gas by absorbing energy from the Sun. Debris is released in this process. Sunlight and the solar wind push on the gas and dust, creating the tails.
- 4 *In which directions do a comet's tails point?* The gas tails of comets point directly away from the Sun; their dust-sized debris tails make arcs pointing away from the Sun.
- 5 *What is a shooting star?* A shooting star is a piece of space debris plunging through Earth's atmosphere—a meteor. It is not a star.

Review Questions

1. A piece of space debris that you pick up from the ground is called a(n):

a. asteroid	d. meteorite
b. meteoroid	e. comet
c. meteor	
2. Space debris that is a roughly equal mix of rock and ice is called a(n):

a. asteroid	d. meteorite
b. comet	e. meteor
c. meteoroid	
3. Which is the rarest type of meteorite found on Earth?

a. iron	c. stony
b. stony-iron	
4. Which part of a comet is solid?

a. nucleus	d. dust tail
b. halo	e. coma
c. gas tail	
5. Suppose you were standing on Pluto. Describe the motions of Charon relative to the horizon. Under what circumstances would you never see Charon?
6. Describe the circumstantial evidence that supports the idea that Pluto is one of a thousand similar icy worlds that once occupied the outer regions of the solar system.

7. What role did Charon play in enabling astronomers to determine Pluto's mass?
8. Why are asteroids, meteoroids, and comets of special interest to astronomers who want to understand the early history of the solar system?
9. Describe the objects in the asteroid belt, including their sizes, orbits, and separation.
10. To test your understanding of the asteroid belt, do Interactive Exercise 8.1 on the assigned Web site. You can print out your results, if required.
11. Why are there many small asteroids but only a few very large ones?
12. Describe the different chemistries of the two tails of a comet.
13. In what directions do comet tails point, and why?
14. What are the Kirkwood gaps, and what causes them?
15. What are the Trojan asteroids, and where are they located?
16. Describe the three main classifications of meteorites. How do astronomers believe that these different types of meteorites originated?
17. Why do astronomers believe that the debris that creates many isolated meteors comes from asteroids, whereas the debris that creates meteor showers is related to comets?
18. To test your understanding of comets, do Interactive Exercise 8.2 on the assigned Web site. You can print out your results, if required.
19. Why is the phrase "dirty snowball" an appropriate characterization of a comet's nucleus?
20. What and where is the Kuiper belt, and how is it related to debris left over from the formation of the solar system?
21. Why do scientists think the Tunguska event was caused by a large meteoroid and not a comet?
22. What evidence in Figure 8-35 supports the labeling of the ion (gas) and dust tails?

Advanced Questions

The answer to the computational problem, which is preceded by an asterisk (), appears at the end of the book.*

23. Would you expect the surfaces of Pluto and Charon to be heavily cratered? Explain.

24. How did the regolith on asteroid Eros form?
25. Why are comets generally brighter after passing perihelion (closest approach to the Sun) than before reaching perihelion?
26. Can you think of another place in the solar system where a phenomenon similar to the Kirkwood gaps in the asteroid belt is likely to exist? Explain your answer.
27. Where on Earth might you find large numbers of stony meteorites that have not been significantly changed by weathering?
- *28. Assuming a constant rate of meteor infall, how much mass has Earth gained in the past 4.6 billion years?

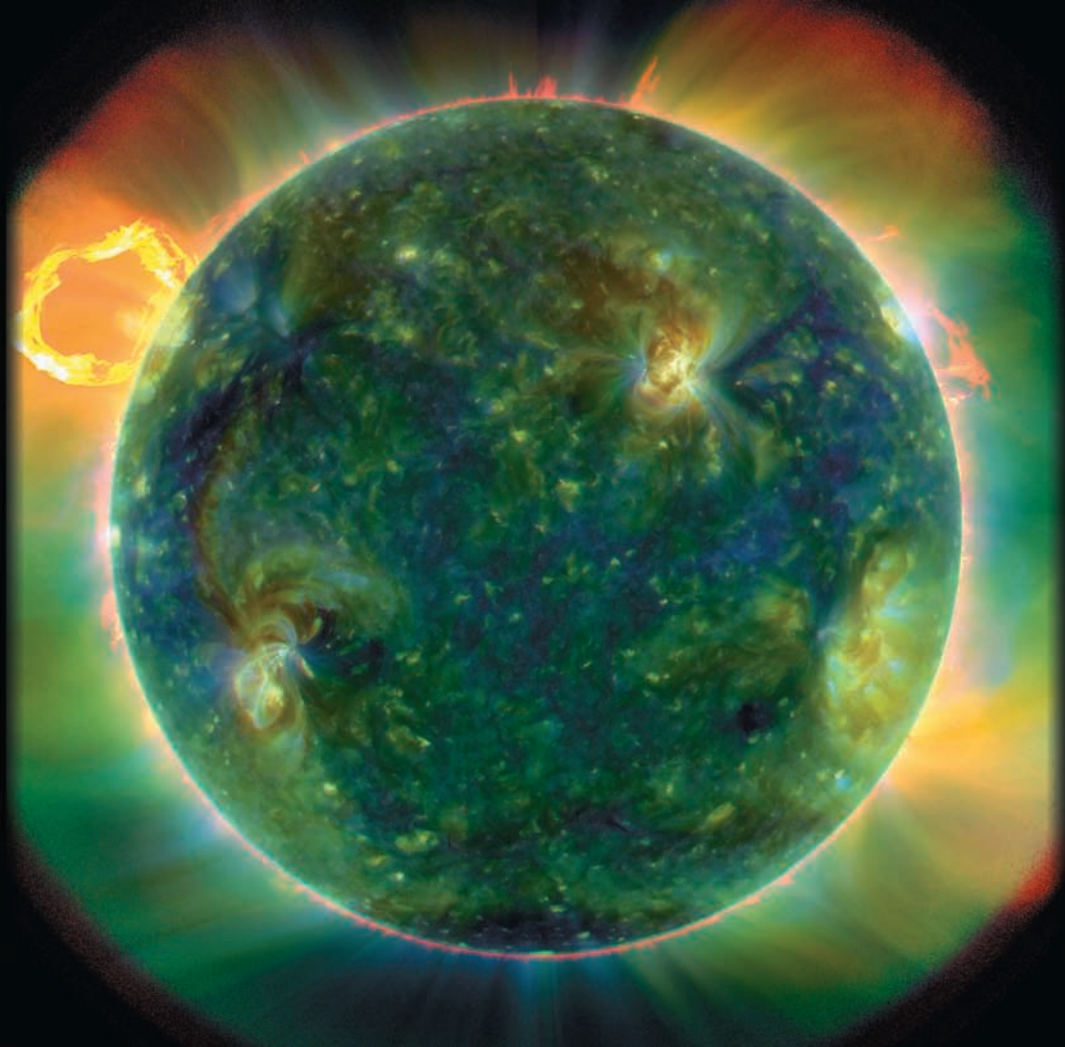
Discussion Questions

29. Suppose it was discovered that the asteroid Hermes had been perturbed in such a way as to put it on a collision course with Earth. Describe what you would do to counter such a catastrophe using present technology.
30. From the abundance of craters on the Moon and Mercury, we know that numerous asteroids and meteoroids struck the inner planets during the very early history of the solar system. Is it reasonable to suppose that numerous comets also pelted the planets 4 to 4½ billion years ago? What effects would such a cometary bombardment have had, especially with regard to the evolution of the primordial atmospheres of the terrestrial planets and oceans on Earth?

Got It?

31. What are the differences between meteors, meteorites, and meteoroids?
32. Where are asteroids found in the solar system?
33. Why is Pluto no longer considered a planet?
34. Are planets the only solar system bodies with moons? If not, what other types of solar system bodies have them?
35. What is the typical separation between asteroids in the asteroid belt?
36. Of the following objects, which are the largest?

- | | |
|---------------|-----------|
| a. asteroids | c. moons |
| b. meteoroids | d. comets |



CHAPTER 9

R I V I U X I G This ultraviolet image of the Sun, taken in 2010 by the Solar Dynamics Observatory, shows a prominence (upper left) in which magnetic fields carry gases above the Sun's surface in a loop that can extend upward hundreds of thousands of kilometers. (NASA/GSFC/AIA)

The Sun: Our Extraordinary Ordinary Star

WHAT DO YOU THINK?

- 1 What percentage of the solar system's mass is in the Sun?
- 2 Does the Sun have a solid and liquid interior like Earth?
- 3 What is the surface of the Sun like?
- 4 Does the Sun rotate? If so, how fast?
- 5 What makes the Sun shine?
- 6 Are matter and energy conserved?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

1 The Sun is the closest star to Earth. The Sun contains as much mass as 333,000 Earths. Indeed, the Sun contains 99.85% of the solar system’s mass. By studying it, astronomers have come to understand how most stars throughout the universe work. As powerful and majestic as it is in our sky, we are going to discover in the next few chapters that the Sun is an ordinary star, neither among the most massive nor among the least massive. Similarly, it is neither among the brightest nor among the dimmest of stars. **Figure 9-1** lists the Sun’s properties.

In this chapter you will discover

- why the Sun is a typical star
- how today’s technology has led to a new understanding of solar phenomena, from sunspots to the powerful ejections of solar matter that sometimes enter our atmosphere and disrupt our technology
- that some features of the Sun generated by its varying magnetic field occur in cycles
- how the Sun generates the energy that makes it shine
- new insights into the nature of matter provided by solar neutrinos

THE SUN: Vital Statistics

Sun symbol:	☉
Mass (1 $M_{\odot}$):	1.989×10^{30} kg = 2.2×10^{27} tons (mass) = 3.33×10^5 Earth masses
Visual diameter (1 $R_{\odot}$):	1.392×10^6 km = 8.65×10^5 mi = 109 Earth diameters
Luminosity (1 $L_{\odot}$):	3.827×10^{26} watts
Mean angular diameter as seen from Earth:	32 arcmin
Rotation periods:	Equatorial: 25 days Polar: 35 days
Mean density:	$1408 \text{ kg/m}^3 = 2373 \text{ lb/yd}^3$
Distances from Earth:	Mean (1 AU): 1.496×10^8 km = 9.30×10^7 mi Maximum: 1.520×10^8 km = 9.44×10^7 mi Minimum: 1.470×10^8 km = 9.13×10^7 mi
Mean light travel time to Earth:	8.32 min
Mean temperatures:	Surface: 5800 K Center: 1.55×10^7 K
Composition (by mass):	71.5% hydrogen (H), 27% helium (He), 1.5% other elements
Composition (by number of atoms):	91.3% hydrogen (H), 8.6% helium (He), 0.1% other elements
Distance from center of Galaxy:	26,000 ly = 8000 pc
Orbital period around center of Galaxy:	230 million years
Orbital speed around center of Galaxy:	220 km/s = 137 mi/s

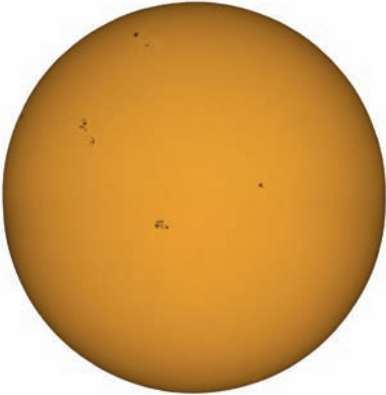




FIGURE 9-1 Our Star, the Sun The Sun emits most of its visible light from a thin layer of gas, called the photosphere, as shown. Although the Sun has no solid or even liquid region, we see the photosphere as its “surface.” Astronomers always

take great care when viewing the Sun by using extremely dark filters or by projecting the Sun’s image onto a screen. (Marcel Clemens/Shutterstock)

THE SUN'S ATMOSPHERE

2 Although astronomers often speak of the solar “surface,” the Sun is so hot that it has neither liquid nor solid matter anywhere inside it. Moving down through the Sun, we continually encounter ever denser and hotter gases.

9-1 The photosphere is the visible layer of the Sun

The Sun appears to have a surface only because most of its visible light comes from one specific gas layer (see Figure 9-1). This region, which is about 400 km (250 mi) thick, is appropriately called the **photosphere** (“sphere of light”). The density of the photosphere’s gas is low by Earth standards, about 0.01% as thick as the air we breathe. The photosphere has a blackbody spectrum (recall Chapter 3) that corresponds to an average temperature of 5800 K (see Figure 3-43).

3 The photosphere is the innermost of the three layers that comprise the Sun’s atmosphere. Because the upper two layers (discussed in the following two sections) are nearly transparent to most wavelengths of visible light, we see through them down to the photosphere. We cannot, however, see through the shimmering gases of the photosphere, so everything below the photosphere is considered to be the Sun’s interior.

As you can see in Figure 9-1, the photosphere appears darkest toward the edge, or **limb**, of the solar disk, a phenomenon called **limb darkening**. This occurs because we see regions of different temperatures at different depths in the photosphere. Here is how it works: Wherever we look on the Sun, we see light through roughly equal amounts of the Sun’s atmosphere. Because the Sun is spherical, the light we see leaving the Sun at different places actually comes from different levels of the photosphere (Figure 9-2). Looking from Earth at the center of the Sun’s disk (see Figure 9-2, bottom black line), we see farther into the Sun’s atmosphere than when we look toward the limb (see Figure 9-2, top black line). The photosphere is hottest and brightest at its base, and so the center of the Sun’s disk, where we see deepest into it, looks brighter and yellower than does its limb.

Under good observing conditions and using a telescope with special dark filters for protection, you can see a blotchy pattern, called *granulation*, on the photosphere (Figure 9-3). The bright **granules** typically measure about 1000 km (620 mi) across and are surrounded by relatively dark boundaries. Time-lapse photography shows that granules form, disappear, and then reform in cycles that last several minutes. At any single moment, several million granules cover the solar surface.

Like liquid in a pot of simmering soup, the gases in granules rise and fall. In Section 3-18 we learned that radial motion of a light source affects the wavelengths of its spectral lines through the Doppler effect (see Figure 3-5).

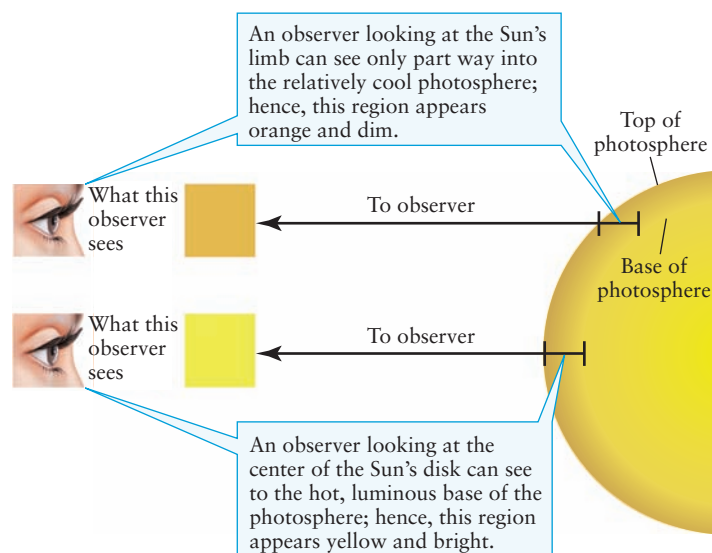


FIGURE 9-2 Limb Darkening The Sun's edge, or limb, appears distinctly darker and more orange than does its center, as seen from Earth (see Figure 9-1). This phenomenon occurs because we look through the same amount of solar atmosphere at all places. As a consequence, we see higher in the Sun's photosphere near its limb than when we look at its central regions. The higher photosphere is cooler and, because it is a blackbody, darker and more orange than the lower, hotter region of the photosphere.

By carefully measuring the Doppler shifts of spectral lines in various parts of individual solar granules, astronomers have determined that hot gases move upward in the center of each granule and cooler gases cascade downward around its edges. This motion is caused by convection. We saw in Section 6-3 that convection moves the continents on Earth and in Section 7-1 that it also helps form the belts and zones on the giant planets. Once the Sun's hot gases arrive at the photosphere, they radiate energy out into space. We see this energy as visible light and other electromagnetic radiation. Upon radiating, the gases cool, spill over the edges of the granules, and plunge back down into the Sun along the boundaries between granules (see Figure 9-3a inset).

Observations indicate that a granule's center is typically 100 K hotter than its edge. According to the Stefan-Boltzmann law (see Section 3-13), hotter regions emit more photons per square meter than do cooler regions. That is why the centers of the granules are brighter than their edges, as you can see in Figure 9-3a.

Focus Question 9-1

Which part of the disk of the Sun appears brighter, the center or the edge? Why?

9-2 The chromosphere is characterized by spikes of gas called spicules

Immediately above the photosphere is a dim layer of less dense stellar gas, called the **chromosphere** (“sphere

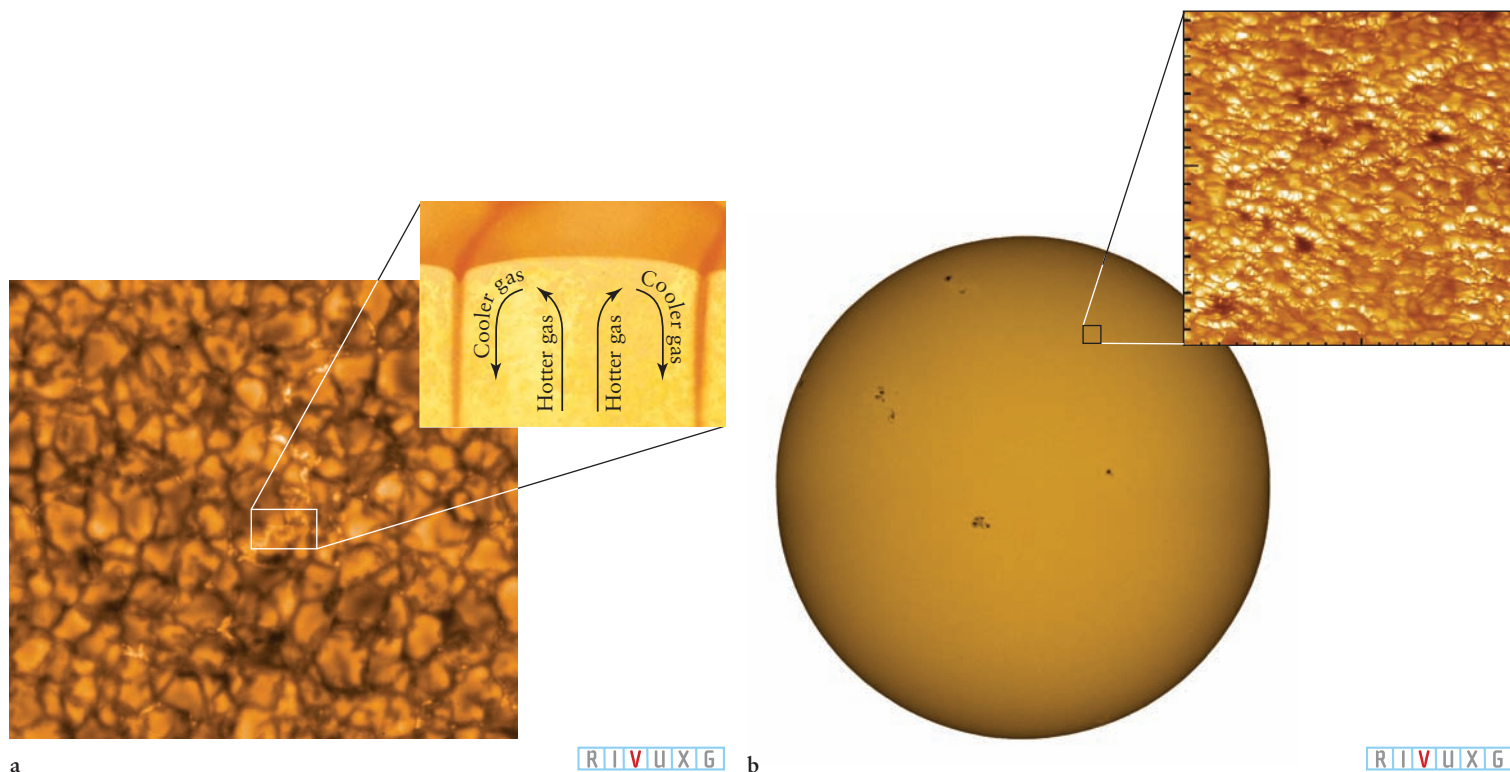


FIGURE 9-3 Solar Granulation (a) High-resolution photographs of the Sun's surface reveal a blotchy pattern, called granulation. Granules, which measure about 1000 km (620 mi) across, are convection cells in the Sun's photosphere. **Inset:** Gas rising upward produces the bright granules. Cooler gas sinks downward along the darker, cooler boundaries between granules. This convective motion transports energy from the Sun's interior outward to the solar atmosphere. (b) At

lower resolution, the Sun's surface appears relatively smooth (the dark regions will be discussed shortly). **Inset:** Viewed near the Sun's limb, granules are seen to bulge upward at their centers as a result of the convection that creates them. (a: Hinode JAXA/NASA/PPARC; b: Marcel Clemens/Shutterstock; inset: G. Scharmer [Swedish 1-meter Solar Telescope, Stockholm University])

of color"). This unfortunate name suggests that it is the layer we normally see, but for centuries the chromosphere was visible only when the photosphere was blocked during a total solar eclipse.

During an eclipse, the chromosphere is visible as a pinkish strip some 2000 km (1200 mi) thick around the edge of the dark Moon. Today, astronomers can also study the chromosphere through filters that pass light with specific wavelengths strongly emitted by it—but not by the photosphere—or through telescopes sensitive to nonvisible wavelengths that the chromosphere emits intensely (Figure 9-4).

High-resolution images of the chromosphere reveal numerous spikes, which are jets of gas called **spicules** (Figure 9-4 and Figure 9-5a). A typical spicule rises for several minutes at the rate of 72,000 km/h (45,000 mi/h) to a height of nearly 10,000 km (Figure 9-5b). Then it collapses and fades away. At any one time, roughly a third of a million spicules cover a few percent of the Sun's chromosphere.

Spicules are generally located on the boundaries of enormous regions of rising and falling chromospheric gas called **supergranules** (Figure 9-5a). A typical supergranule has a diameter slightly larger than Earth's and contains about 900 granules.

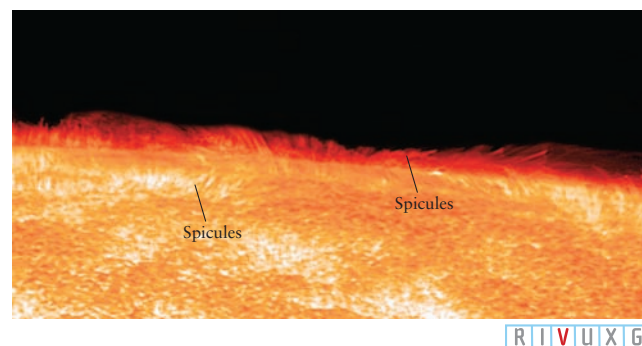
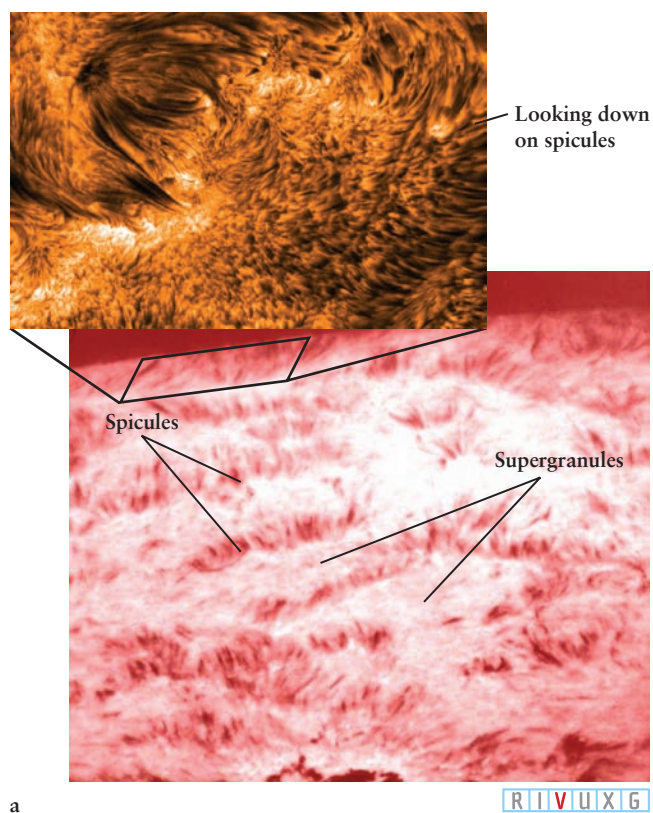


FIGURE 9-4 The Chromosphere This photograph of the chromosphere was taken by the *Hinode* (Japanese for “sun-rise”) satellite. The dark bumps are the tops of granules, and the light regions are hotter gases in spicules. The spicules on the edge or limb of the Sun give a sense of the height of these gas jets. (Hinode JAXA/NASA)

Focus Question 9-2

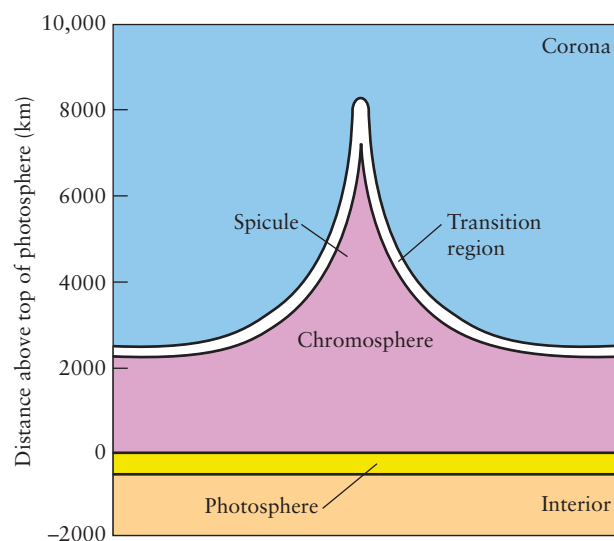
Which part of the Sun do we normally see?



a

R I V U X G

FIGURE 9-5 Spicules and Supergranules (a) Spicules appear in this photograph of the Sun's chromosphere. The Sun appears rose-colored in this image because the photo was taken through a H_α filter that passes red light from hydrogen and effectively blocks most of the photosphere's light. Surrounded by spicules, supergranules are regions of rising and falling gas in the chromosphere. Each supergranule spans hundreds of granules in the photosphere below. **Inset:** A view of spicules from above. (b) The spicules are jets of gas that surge upward into the Sun's outer atmosphere. This schematic diagram shows a



b

spicule and its relationship to the solar atmosphere's layers. The photosphere is about 400 km (250 mi) thick. The chromosphere above it extends to an altitude of about 2000 km (1200 mi), with spicules jutting up to nearly 10,000 km (6200 mi) above the photosphere. The outermost layer, the corona (discussed in Section 9-3), extends millions of kilometers above the photosphere. (a: NASA; inset: Swedish Solar Telescope/Royal Swedish Academy of Sciences, La Palma, Spain, by Bert De Pontieu, Lockheed Martin Solar and Astrophysics Lab)

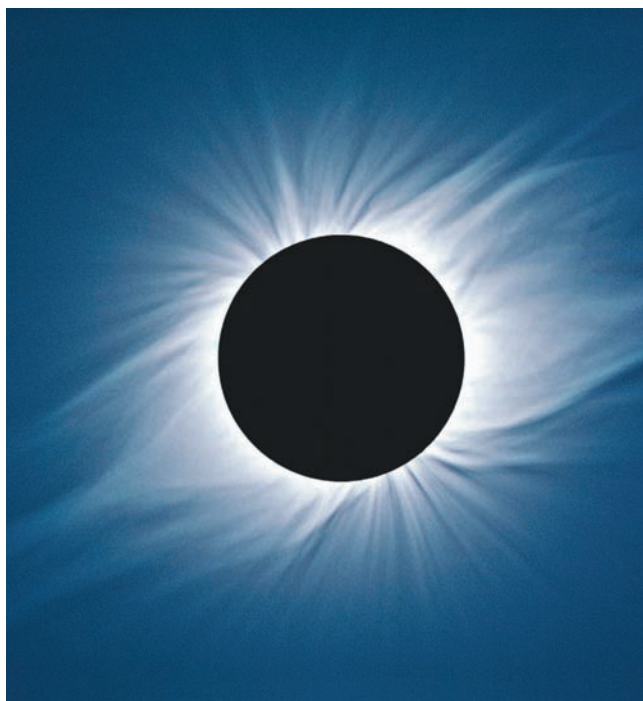
9-3 The outermost layer of the Sun's atmosphere, the corona, is exceptionally hot

It seems plausible that the temperature should decrease as you rise through the Sun's atmosphere. After all, by moving upward, you move farther from the Sun's internal heat source. Indeed, starting in the photosphere at 5800 K, the temperature drops to around 4000 K in the lower chromosphere. Surprisingly, however, the temperature then begins to increase as you ascend, reaching about 10,000 K at the top of the chromosphere. The outermost region of the Sun's atmosphere, the **corona**, extends several million kilometers from the top of the chromosphere (Figure 9-6). Between the chromosphere and corona is a **transition zone** in which the temperature skyrockets to about 1 million K (see Figure 9-6c).

This unexpected increase in temperature was discovered around 1940 as a result of the high temperature's effect on the spectrum of the Sun's corona. The hotter

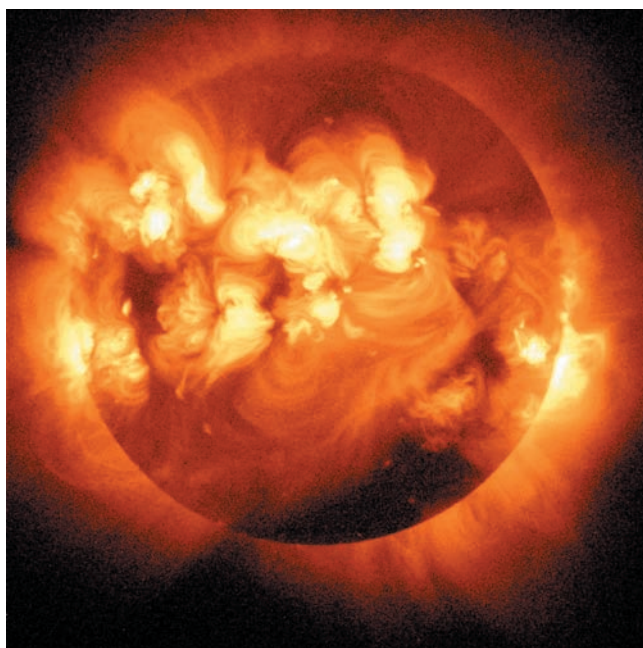
a gas is, the more it is ionized (electrons are stripped off the atoms). Astronomers discovered that the corona contains the emission lines of several highly ionized (therefore, very hot) elements. For example, a prominent green line caused by the presence of Fe XIV, an iron atom stripped of 13 electrons, appears in the coronal spectrum. (Recall from Section 3-16 that Fe I is a neutral iron atom.) It is now known that coronal temperatures are typically in the range of 1 million to 2 million K, although some regions are even hotter.

Although its temperature is extremely high, the density of gas in the corona is very low, about 10 trillion times less dense than the air at sea level on Earth. The low density partly accounts for the dimness of the corona, which otherwise would outshine the photosphere. Astronomers have observations suggesting two mechanisms that heat the corona. First, energy is apparently carried aloft and released there by the Sun's complex magnetic fields, discussed in Sections 9-5 and 9-6. Second, jets of very hot gas, heated by a mechanism still



a

R I V U X G



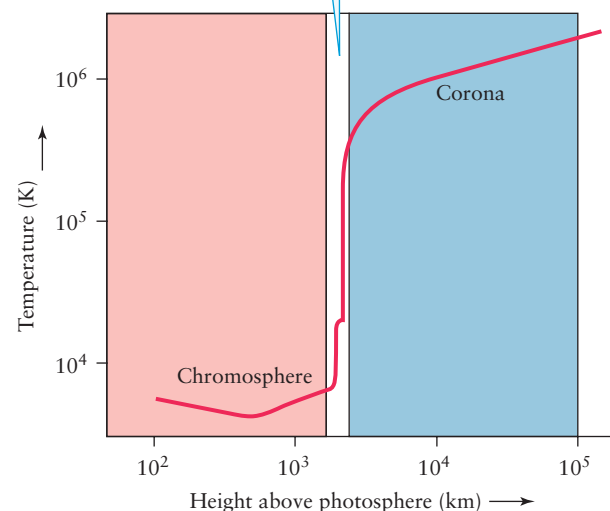
b

R I V U X G

under study, are sent up from the chromosphere into the corona.

The total amount of visible light that we receive from the solar corona is comparable to the brightness of the full Moon—or only about one-millionth as bright as the photosphere. As with the chromosphere, the corona can be seen only when the photosphere is blocked out or through special filters or at nonvisible wavelengths (such as ultraviolet and X-ray), at which the corona

In this narrow transition region between the chromosphere and corona, the temperature rises abruptly by about a factor of 100.



c

FIGURE 9-6 The Solar Corona (a) This visible-light photograph was taken during the total solar eclipse of July 11, 1991. Numerous streamers are visible, extending millions of kilometers above the solar surface. (b) This X-ray image of the Sun's corona, taken by the *Yohkoh* satellite in 1999, provides hints of the complex activity taking place on and in the Sun. The million-degree gases in the corona emit the X rays visible here. (c) This graph shows how temperature varies with altitude in the Sun's chromosphere and corona and in the transition region between them (white). Note that both the height and temperature scales are nonlinear. (a: Jerry Lodriguss/Science Source; b: JAXA/NASA; c: data from A. Gabriel)

is especially bright compared to the photosphere and chromosphere.

The photosphere is blocked naturally during a total eclipse or artificially with a specially designed telescope, called a *coronagraph*. Figure 9-6a is a photograph of the corona taken during a total eclipse (see also Figure 1-26). Figure 9-6b shows a stunning X-ray image of the corona.

Just as Earth's gravity prevents most of our atmosphere from escaping into space, so, too, does the Sun's gravity keep most of its outer layers from leaving. However, some of the gas is moving fast enough—around a million kilometers per hour—to escape the Sun's gravity forever and race into space. A portion of this gas comes directly from the corona, with the help of the Sun's magnetic fields, while some of it is funneled out from below the corona, also by magnetic fields. This outflow of particles is called the **solar wind**. (We saw its effects on Earth's magnetosphere in Section 6-5 and on comets in Section 8-7.) The *heliosphere* is a bubble in space

created by the solar wind and solar magnetic field that contains the Sun and planets. The heliosphere extends from the Sun outward approximately 115 AU. The solar wind and the solar magnetic field that it drags out with it prevent most of the gases expelled in our direction by other stars from entering our solar system. During the years 2008–2010, the solar wind was the weakest it has been in recorded history. As a result, the heliosphere allowed in more particles than usual from outside the solar system, called *galactic cosmic rays*.

In 2008, using data from the two venerable *Voyager* spacecraft that were launched in 1977 and that are now in the process of leaving the solar system, astronomers found that the heliosphere is not spherical. This discovery happened because the two spacecraft were sent out in different directions (and launched 16 days apart), and *Voyager 2* reached the inner edge of the heliosphere closer to the Sun than did *Voyager 1*. In August 2012, the *Voyager 1* spacecraft departed from the heliosphere. *Voyager 2* followed shortly. They are the first human-made objects to enter interstellar space.

The information from the *Voyager* spacecraft provides only two data points about the location of the boundary of the heliosphere. To get a complete map of this boundary, NASA launched the Interstellar Boundary Explorer (IBEX) in 2008. In 2013, NASA's IBEX satellite discovered that the solar system has a tail of solar wind gases that extends out beyond the heliosphere into interstellar space. Guided by the Sun's magnetic field, this tail is believed to dissipate by mixing with other interstellar gas.

The Sun ejects around 9×10^8 kg (a million tons) of matter each second as the solar wind. Even at this rate of emission, the mass loss due to the solar wind will amount to only a few tenths of a percent of the Sun's total mass throughout its lifetime. The solar wind particles reach speeds up to 2.9×10^6 km/h (1.8×10^6 mi/h). The wind achieves these high speeds, in part, by being accelerated by the Sun's magnetic field, discussed in Section 9-5. By the time the solar wind travels 1 AU and arrives in the vicinity of Earth, it has spread out so much that there are typically just nine particles per cubic centimeter (cm^3), and they are moving around 1.1 to 1.4×10^6 km/h (6.8 to 8.5×10^5 mi/h). For comparison, there are 6×10^{19} molecules/ cm^3 in the air we breathe, and the winds at Earth's surface are moving less than 240 km/h (150 mi/h).

Astronomers think that the Sun's surface chemical composition today is nearly identical to the composition of the solar nebula (see Section 4-2). Although electrons and hydrogen and helium nuclei comprise 99.9% of the solar wind, silicon, sulfur, calcium, chromium, nickel, neon, and argon ions have also been detected in it. To better understand the chemistry of the early solar system, the *Genesis* spacecraft was launched to collect pristine solar wind particles far outside Earth's magne-

tosphere (see Section 6-5) and to return them to Earth. By studying the amounts of the different chemicals in the solar wind, astronomers hope to better understand how the Sun works, as well as to learn the chemistries of the earlier generation of stars that exploded and whose gases eventually recondensed and formed the Sun and the other objects in the solar system. While attempting to land in 2004, *Genesis*'s parachute failed and the spacecraft crashed into Earth at 320 km/h (about 200 mi/h) (see Section 8-7). Nevertheless, the collected samples have been salvaged, and as mentioned in Section 8-7, they reveal that the amounts of each isotope of nitrogen and oxygen (atoms of nitrogen and oxygen that have different numbers of neutrons in their nuclei) in the solar wind and Earth's atmosphere are not identical, leading to new questions about the evolution of the early solar system.

Another important result from the *Genesis* mission concerns the chemistry of neon in the Sun. The amounts of different neon isotopes in the Moon rocks brought back by Apollo astronauts suggested that the Sun was more active (hotter) in the past than it is today. However, the amounts of neon isotopes collected by *Genesis* were different from those in the Moon rocks. Astronomers deduce that the neon at different depths of the Moon was chemically altered by different amounts due to the radiation from space over billions of years. Taking this into account, they conclude that the Sun has maintained a more nearly constant temperature than previously believed.

Focus Question 9-3

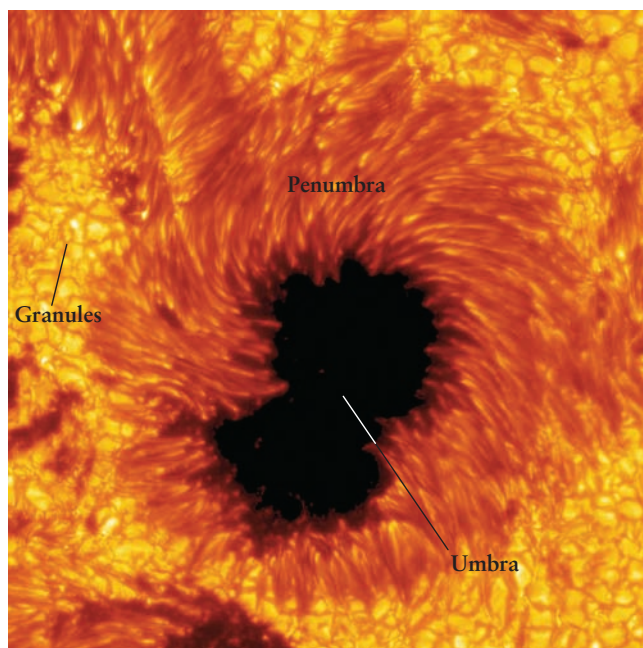
Which of the Sun's three atmospheric layers is coolest? Which is densest?

THE ACTIVE SUN

Granules, supergranules, spicules, and the solar wind occur continuously. But the Sun's atmosphere is also periodically disrupted by magnetic fields that stir things up, creating a group of phenomena known collectively as the *active Sun*. The Sun's most obvious transient features are **sunspots**, regions of the photosphere that appear dark because they are cooler than the rest of the Sun's lower atmosphere. Sometimes sunspots occur in isolation (**Figure 9-7a**), but often they arise in clusters, called *sunspot groups* (**Figure 9-7b**).

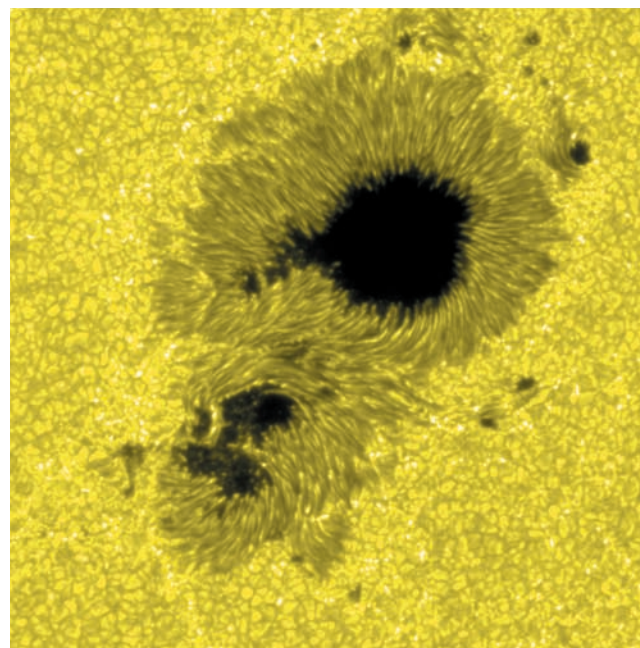
9-4 Sunspots reveal the solar cycle and the Sun's rotation

Like other transient features of the active Sun, the average number and location of sunspots vary in fairly predictable cycles because of the changes in magnetic fields that produce them (as we will explore in Section 9-5). The sunspot cycle was discovered in 1843 by the German astronomer Samuel Schwabe. As shown in **Figure 9-8a**, the average sunspot cycle lasts approximately 11 years. Within an 11-year cycle, the largest



a

FIGURE 9-7 Sunspots (a) This dark region on the Sun is a typical isolated sunspot. Granulation is visible in the surrounding, undisturbed photosphere. (b) This high-resolution photograph



b

shows a sunspot group in which several sunspots in a large group overlap. (a: G. Scharmer (Swedish 1-meter Solar Telescope, Stockholm University); b: JAXA/NASA/Hinode)

numbers of sunspots are seen during the **sunspot maximum** of that cycle (Figure 9-8b). Sunspot maxima occurred most recently in 1979, 1989, 2001, and 2013. During a **sunspot minimum**, the Sun is often devoid of sunspots, as it was in 1986, 1997, and 2008 (Figure 9-8c). The next minimum in the cycle is projected to occur in 2019.

A typical sunspot is 10,000 km (6200 mi) across and lasts between a few hours and a few months. Each sunspot has two parts: a dark, central region, called the *umbra*, and a brighter ring that surrounds the umbra, called the *penumbra* (see Figure 9-7). Although these are the same names as the blocked areas of an eclipse (see Section 1-13), we will see in the next section that their causes are completely different—another good example of how certain astronomical terms can have different meanings in different contexts.

Seen without the surrounding brilliant granules that outshine it, a sunspot's umbra appears red and its penumbra is orange. According to Wien's law (see Section 3-13), these colors indicate that the umbra is typically 4300 K and the penumbra is 5000 K, both cooler than the normal photosphere.

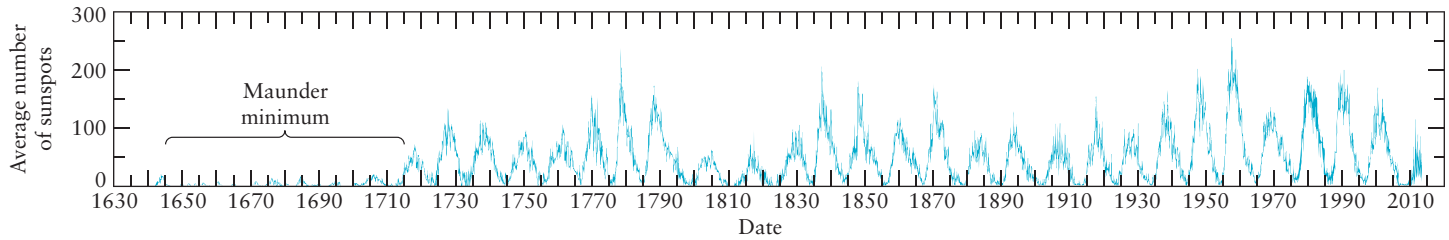
On rare occasions, a sunspot group is so large that it can be seen with the unaided eye. Chinese astronomers recorded such sightings 2000 years ago, and a huge sunspot group visible to the naked eye was seen in 2001. *Always use special dark filters or other means to protect*

your eyes when viewing the Sun. Looking directly at the Sun for more than a few moments causes eye damage! Of course, a telescope gives a much better view, so it was not until Galileo used one that anyone examined sunspots in detail.

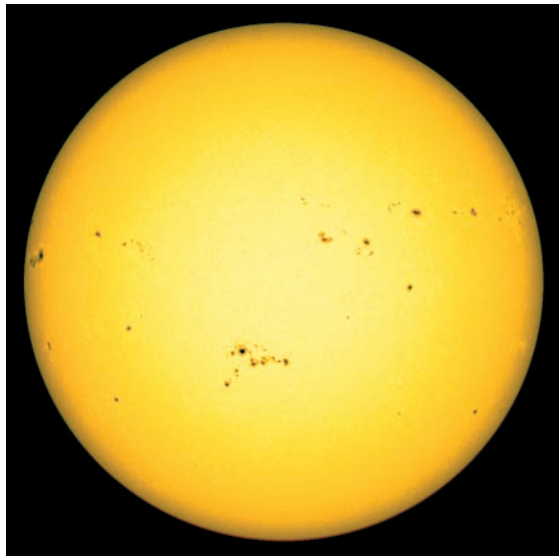
4 By following sunspots as they moved across the solar disk (**Figure 9-9**), Galileo discovered that *the Sun rotates once in about 4 weeks*. He did this by observing the Sun indirectly, using the technique developed by his protégé Benedetto Castelli of projecting the telescope image of the Sun on a screen. (It is unlikely that Galileo's blindness in old age was caused by looking at the Sun. Rather, it is believed that his blindness came as a result of medical issues, probably glaucoma and cataracts.)

Insight Into Science

Perception versus Reality Our senses (and often our technology) are limited, so we must be careful about interpreting what we perceive. For example, the sunspots in Figure 9-7 certainly look like black spots. However, they appear black only in contrast to the bright light around them. As we will discuss shortly, sunspots are actually red and orange.

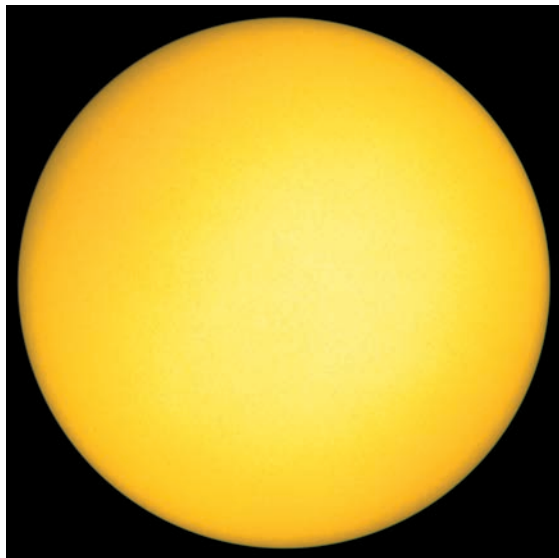


a



b

R I V U X G

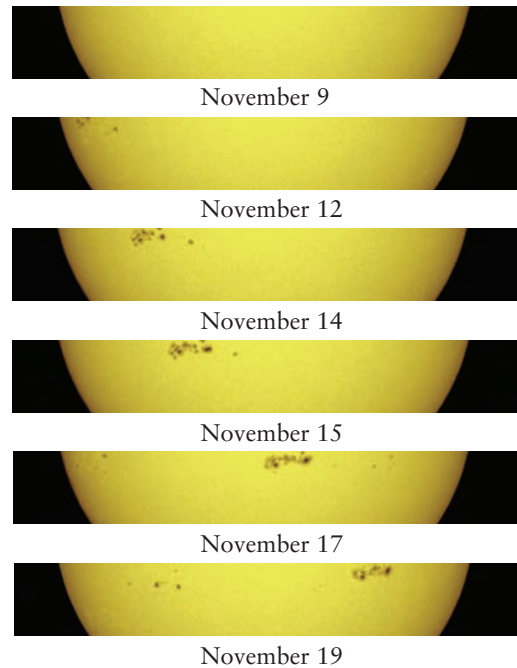


c

R I V U X G

Because a typical sunspot group lasts about 2 months, it can be followed for two solar rotations. Sunspot activity also reveals that like the giant planets, different latitudes of the Sun rotate at different rates, a phenomenon called **differential rotation**: The equatorial regions rotate more rapidly than the polar regions. A sunspot near the solar equator takes 25 days to go once around the Sun, whereas a sunspot at 30° north or south of the equator takes about 27 days. The rotation period at 75° north or south of the equator is about 33 days, and near the poles it is as long as 35 days.

The average latitude at which new sunspots appear changes throughout the sunspot cycle. At the beginning of each cycle, the sunspots appear mostly at about 30° north and south latitudes. Those sunspots that form later in the cycle typically occur closer to the equator.



November 19

R I V U X G

FIGURE 9-8 The Sunspot Cycle (a) The number of sunspots on the Sun varies with a period of about 11 years. The most recent sunspot maximum occurred in 2013, and the most recent sunspot minimum began in 2007 (and continued through 2010). (b) The active Sun has many sunspots (this photo was taken in 1979). (c) The Sun has many fewer sunspots when it is not active (this photo, showing a time with no sunspots, was taken in 1986). (a: Hathaway/NASA/MSFC; b and c: SOHO [ESA & NASA])

FIGURE 9-9 The Sun's Rotation This series of photographs taken in 1999 shows the same sunspot group over one-third of a solar rotation. Note how the sunspot groups changed over this time. By observing a group of sunspots from one day to the next in this same manner, Galileo found that the Sun rotates once in about 4 weeks. Sunspot activity also reveals the Sun's differential rotation: The equatorial regions rotate faster than the polar regions. (Carnegie Observatories)

DAILY SUNSPOT AREA AVERAGED OVER INDIVIDUAL SOLAR ROTATIONS

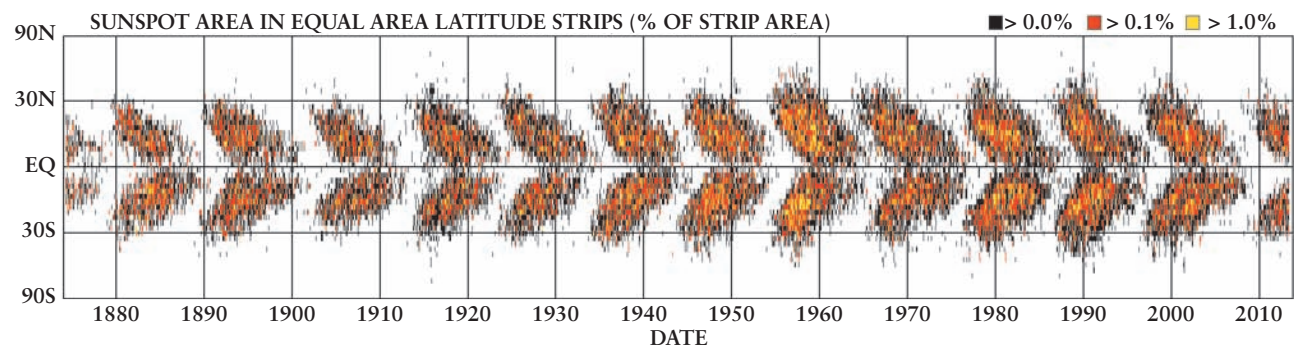


FIGURE 9-10 Locations of Sunspots Throughout the Sunspot Cycle This “butterfly” diagram of sunspot locations shows that sunspots occur at changing latitudes throughout each cycle.

From most common locations to least, the diagram is color-coded yellow, orange, and black. (Hathaway/NASA/MSFC)

The “butterfly diagram” presented in **Figure 9-10** shows that the range of latitudes on which sunspots are located and the average latitude at which they are found vary at the same 11-year rate as does the number of sunspots. Satellite measurements reveal that the Sun emits about 0.1% more energy at the peak of the sunspot cycle than at its minimum.

9-5 The Sun’s magnetic fields create sunspots

In 1908, the American astronomer George Ellery Hale discovered that sunspots are directly linked to intense magnetic fields on the Sun. When Hale focused a spectroscope on sunlight coming from a sunspot (**Figure 9-11a**), he found that each spectral line in the normal solar spectrum is flanked by additional, closely spaced spectral lines not usually observed (**Figure 9-11b**). This “splitting” of

Focus Question 9-4
Which area of the photosphere is hottest, a sunspot’s umbra, its penumbra, or a granule?

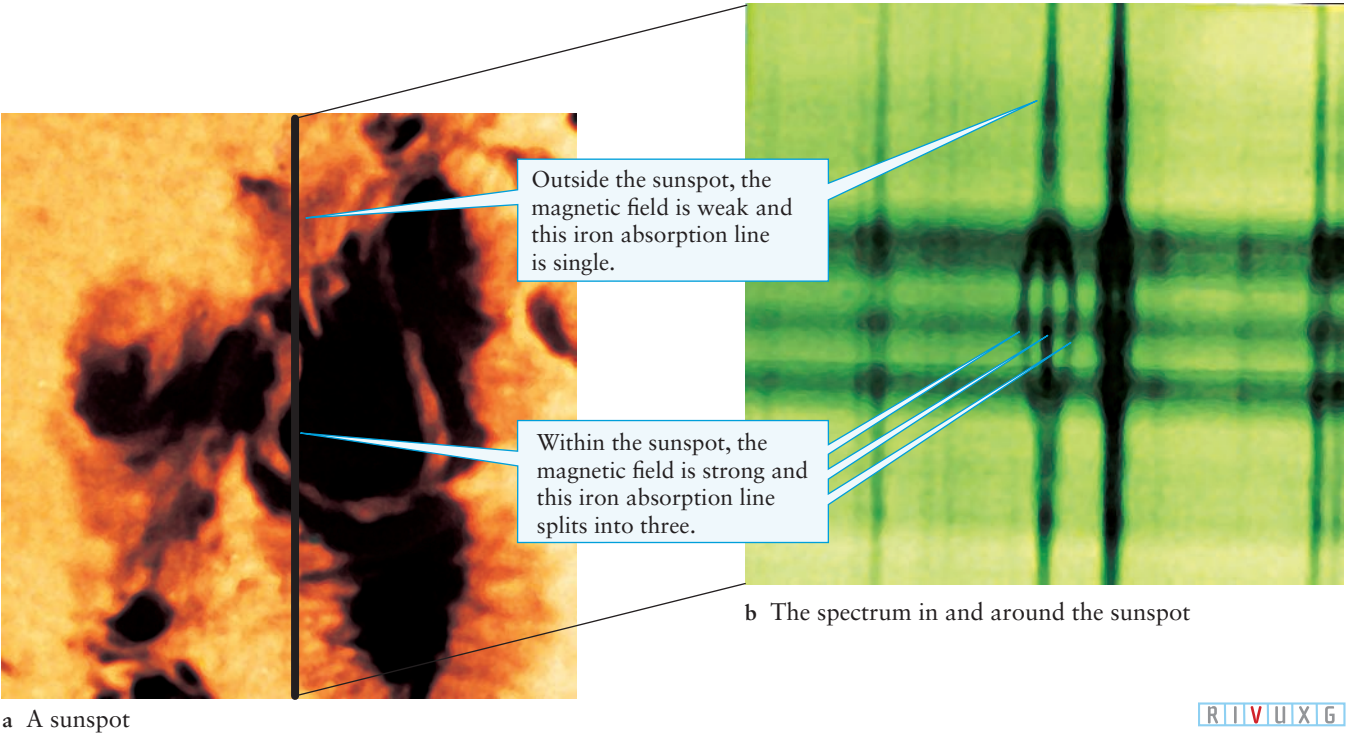


FIGURE 9-11 Zeeman Splitting by a Sunspot’s Magnetic Field (a) The black line drawn across the sunspot indicates the location toward which the slit of the spectroscope was aimed. (b) In the resulting spectrogram, one line in the middle of the normal solar spectrum is split into three components by the

Sun’s magnetic field. The amount of splitting between the three lines is used to determine the magnetic field’s strength. Typical sunspots have magnetic fields some 5000 times stronger than Earth’s magnetic field. (a and b: NSO/AURA/NSF)

a single spectral line into two or more lines is called the **Zeeman effect**. The Dutch physicist Pieter Zeeman, who first observed it in the laboratory in 1896, showed that an intense magnetic field splits the spectral lines of a light source inside the field. The more intense the magnetic field, the more the split lines are separated.

Sunspots are areas where concentrated north or south magnetic fields project through the hot gases of the photosphere. But how do these fields create the spots? The answer lies in the interaction between the magnetic fields and the photosphere's gases. Because of the photosphere's high temperature, many atoms in it are *ionized*: One or more of their electrons have been stripped off by high-energy photons there. As a result, the photosphere is a mixture of electrically charged ions and electrons, called a **plasma**.

Plasmas are extremely good conductors of electricity and are repelled from regions of high magnetic field. Therefore, the magnetic field protruding through the photosphere prevents hot, ionized gases inside the Sun from convecting to the surface as they normally do. Thus, such regions of the photosphere are left relatively devoid of hot gas and are therefore cooler and darker than the surrounding solar surface. We see these darker, cooler regions as sunspots.

The fact that magnetic fields create sunspots presented a problem for astronomers studying the Sun. If you have ever played with toy magnets, you may have discovered that two north poles or two south poles repel each other. How is it then that bundles of same-pole magnetic fields can stay together where they pierce the Sun's photosphere? In other words, why don't they push each other apart and thereby create many smaller spots on the Sun? The answer lies under the Sun's visible layer.

To learn more about the Sun's interior, astronomers record its vibrations—a study called **helioseismology**—just as geologists use earthquakes to study Earth's interior structure. Although there are no true sunquakes, the Sun does vibrate in a variety of ways, somewhat like a vibrating drum head, which, when struck in different places, makes different sounds due to different modes of vibration. Solar vibrations, first noted in 1960, can be detected by using sensitive Doppler shift measurements. Astronomers have observed, for example, that portions of the Sun's surface move up and down by about 10 km (6 mi) every 5 minutes (**Figure 9-12a**).

Slower vibrations with periods ranging from 20 minutes to nearly an hour were discovered in the 1970s. Oscillations lasting several days have since been detected, as have pulses 16 months long. One important discovery

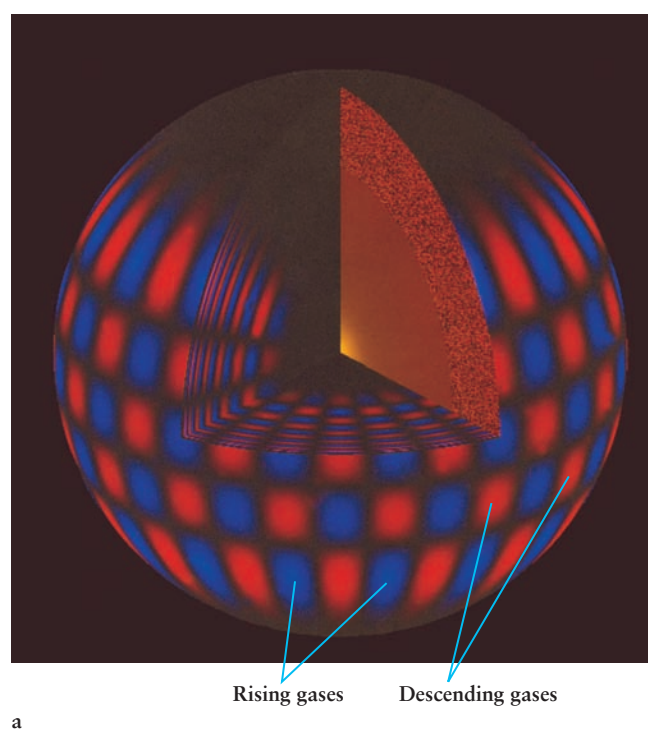
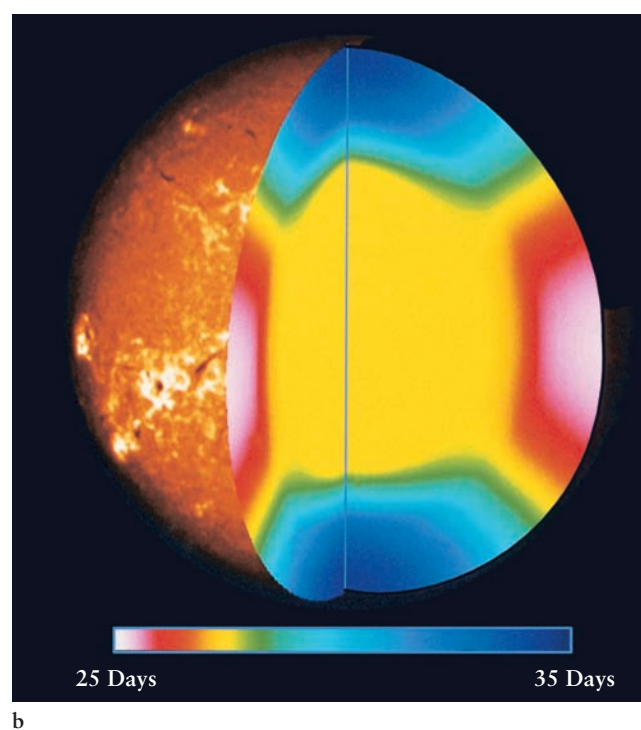


FIGURE 9-12 Helioseismology (a) This computer-generated image shows one of the myriad ways in which the Sun vibrates because of sound waves resonating in its interior. The regions that are moving outward are *blue*; those moving inward are *red*. The cutaway shows how deep these oscillations are believed to extend. (b) This cutaway picture of the Sun shows how the rate of solar rotation varies with depth and latitude. *Red* and *yellow* denote faster-than-average motion; *blue* regions move



more slowly than average. The pattern of differential surface rotation, which varies from 25 days at the equator to 35 days near the poles, persists at least 19,000 km (11,800 mi) down into the Sun's convective layer. Sunspots preferentially occur on the boundaries between different rotating regions. Earthlike jet streams and other wind patterns have also been discovered in the Sun's atmosphere. (a: NSO/AURA/NSF; b: Big Bear Solar Observatory/New Jersey Institute of Technology)

from helioseismology is that deep inside, the Sun rotates like a rigid body, rather than with the differential rotation we see on its surface (Figure 9-12b).

While observing the vibrations of the Sun around sunspots, astronomers in 2001 apparently found the answer to the question of how sunspots persist for months despite the repulsion of the magnetic fields inside them. They discovered that below the photosphere, the gases that surround each sunspot whip around like hurricanes as large across as Earth's diameter. The circulation of charged gases around the magnetic fields holds sunspots in place.

Recall from Section 6-5 that magnetic field lines form complete loops, with each magnetic field having a north pole and a south pole. Likewise, when the Sun's magnetic field emerges through one sunspot or sunspot group, it forms a loop that reenters the Sun at another sunspot or sunspot group (Figure 9-13). We associate the names *north pole* or *south pole* with each sunspot or sunspot group, depending on whether the magnetic field there is pointing outward (a south pole) or pointing inward (a north pole). Sunspots and sunspot groups are connected in pairs—one where the magnetic field points out of the Sun, the other where it points into the Sun.

During his observations, Hale found that on one hemisphere of the Sun, sunspots with a magnetic north pole always come into view before the corresponding sunspots with a magnetic south pole. At the same time on the other hemisphere, the order is reversed (see Figure 9-13). Hale also found that this pattern reverses itself about every 11 years. The hemisphere where north magnetic poles come first during one 11-year cycle has south magnetic poles coming first during the next. Astronomers, therefore, speak of the 22-year **solar cycle**, the time it takes solar magnetic fields to return to their original orientation. The Sun is not alone in having its magnetic field reverse. In 2007, astronomers observed the flipping of the magnetic fields of the star Tau Boötis, located 50.9 ly from Earth.

Based on these observations, we can now explain the Sun's magnetic field. This field is created as a result of the Sun's rotation and the resulting motion of the ionized particles found throughout it. This theory was proposed in 1960 by another American astronomer, Horace Babcock, as the **magnetic dynamo** model, in an effort to explain the 22-year solar cycle. When the Sun is quiet, its magnetic field lies just below the surface in the highly

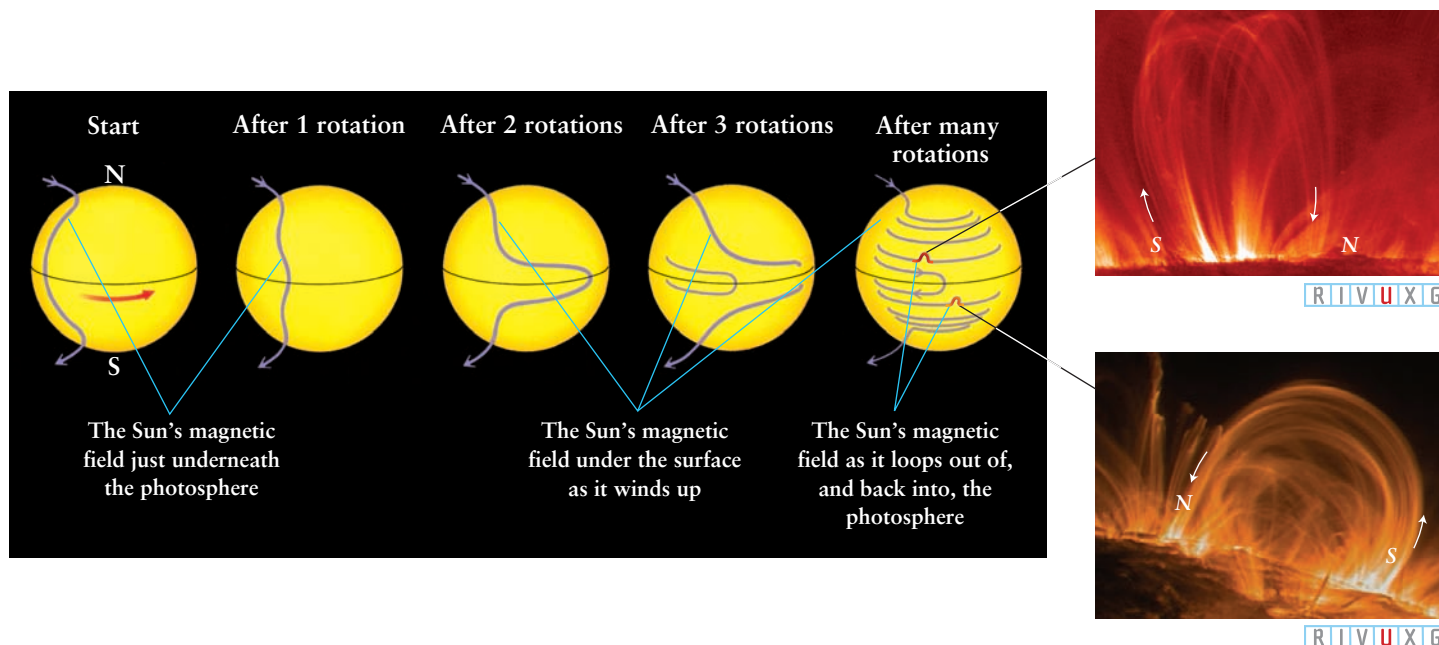


FIGURE 9-13 Babcock's Magnetic Dynamo In a plausible partial explanation for the sunspot cycle, differential rotation wraps a magnetic field around the Sun, just under its surface. Convection under the photosphere tangles the field, which becomes buoyant and rises through the photosphere, creating sunspots and sunspot groups. **Insets:** In each group, the sunspot that appears first as the Sun rotates (on the right side of each loop as you view them) has the same polarity as the Sun's magnetic pole in that hemisphere (*N* in the upper hemisphere and *S* in the lower hemisphere). To understand why, note that magnetic fields make complete loops: If a north pole enters the Sun, it leads to a south pole emerging, and vice versa. Follow the magnetic field in the northern hemisphere from where it

enters the Sun (near the top) in counterclockwise loops until you encounter the bump in that hemisphere. Where the field emerges is the bump is a south pole and where it reenters the Sun is a north pole, as shown. Conversely, following the fields in the southern hemisphere, starting at the south pole, you encounter the southern bump from the left, emerging as a north pole and reentering the Sun as a south pole, as drawn. The Sun's magnetic fields are revealed by the radiation emitted from the gas they trap. These ultraviolet images show coronal loops up to 160,000 km (100,000 mi) high, with gases moving along the magnetic field lines at speeds of 100 km/s (60 mi/s). (Top inset: TRACE, Stanford-Lockheed Institute for Space Research, and NASA Small Explorer Program; bottom inset: NASA)

conducting plasma located there, unlike Earth's field, which passes through our planet's center.

As shown in Figure 9-13, the Sun's differential rotation causes its magnetic field to become increasingly stretched. Like expanding a rubber band, stretching the field this way causes it to store energy. At the same time, the magnetic field becomes tangled like hair that has not been combed in a while as convection of the gases under the photosphere also causes the fields to rise and fall. Unlike a rubber band, however, magnetic field lines cannot break to release the energy stored in them. Rather, the fields must untangle themselves.

This untangling process begins as the mixed-up regions of magnetic field trap gases. This gas is compressed and heated. Therefore, it expands, and like bubbles in a glass of soda, becomes buoyant and floats up through the solar surface, carrying the magnetic fields with it. Sunspots, filled with magnetic fields typically 5000 times stronger than Earth's magnetic field, form where loops or tangles of solar magnetic field leave and reenter the Sun. The gas bottled up in the loops of magnetic field eventually leaks out, and the fields untangle, interact with other parts of the Sun's magnetic field, and gradually settle back under the photosphere, at which point the sunspots associated with them disappear. As noted earlier, this process of magnetic fields piercing the Sun's surface begins at moderate latitudes, and over the next 11 years it occurs closer and closer to the Sun's equator (see Figure 9-10).

Because of how these fields interact and vanish, every 11 years the Sun's entire magnetic field is reversed—the Sun's north magnetic pole becomes its south magnetic pole, and vice versa. This occurs at the peak of each sunspot cycle. After another 11-year cycle, the field is back to its original orientation. This is why the solar cycle is 22 years long. The most recent reversal of the Sun's magnetic field occurred in 2013.

Notable irregularities occur in the solar cycle. For example, the overall reversal of the Sun's magnetic field is often piecemeal and haphazard. More intriguing still is the strong historical evidence that all traces of sunspots and the sunspot cycle have vanished for decades at a time. For example, in 1893 the British astronomer E. Walter Maunder used historical observations to conclude that virtually no sunspots occurred from 1645 through 1715 (see Figure 9-8a). This period, called the *Maunder minimum*, coincides with a period of cold so extreme in Europe that it was called the *Little Ice Age*. At the same time, western North America was subject to severe drought.

Similar sunspot-free periods apparently occurred at irregular intervals in earlier times as well. Conversely, periods of increased sunspot activity in the eleventh and twelfth centuries coincided with periods of warmer-than-average temperatures. Interestingly, the minimum that began in 2007 went on longer than any minimum since the one of 1913. Furthermore, the peak of the present sunspot cycle in 2013 had the fewest sunspots in 200 years.

Whether these variations from the norm have or will have any bearing on Earth's global weather remains to be seen. Indeed, the question of whether there is a cause-and-effect link between the number of sunspots and periods of extreme temperatures on Earth is still open.

Insight Into Science

Cause and Effect While extremes of sunspot activity often occur at the same times as weather extremes on Earth, models of these phenomena must also take into account the possibility of purely terrestrial causes for these temperature changes. Remember, just because one event follows another does not mean that the first *causes* the second.

9-6 Solar magnetic fields also create other atmospheric phenomena

Figure 9-14 shows the active Sun's chromosphere and lower corona. The bright areas in this photograph are called **plages** (pronounced *plahzh*, from the French word for “beaches”). By studying the light emitted by calcium or hydrogen atoms in the plages, we know that

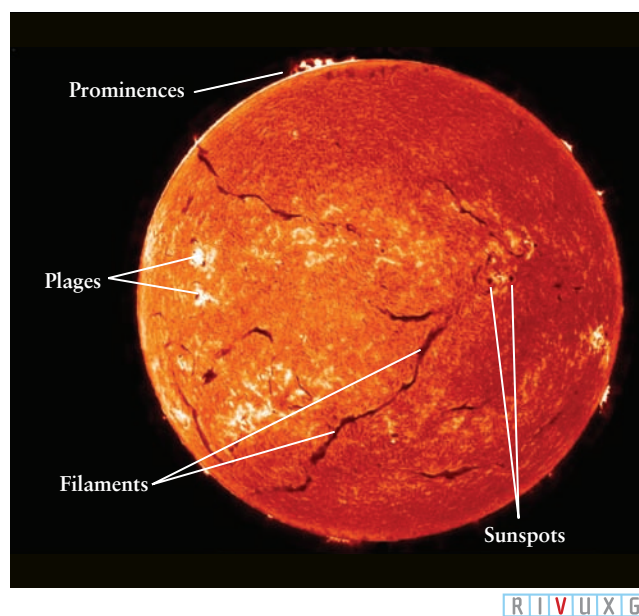
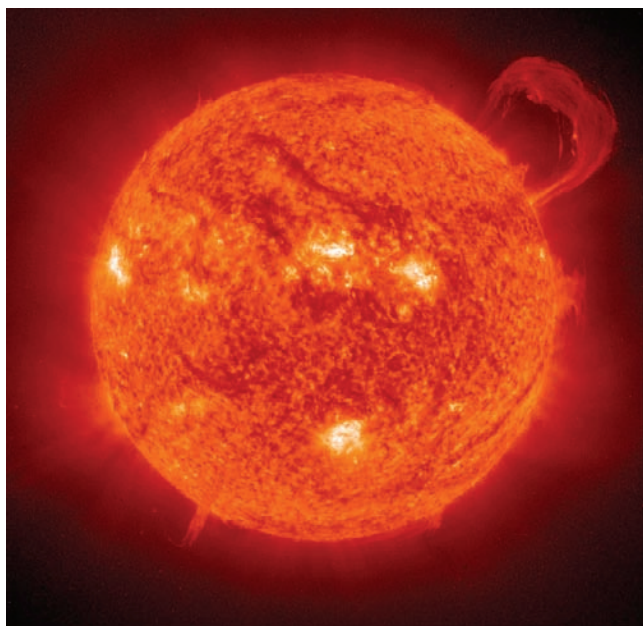


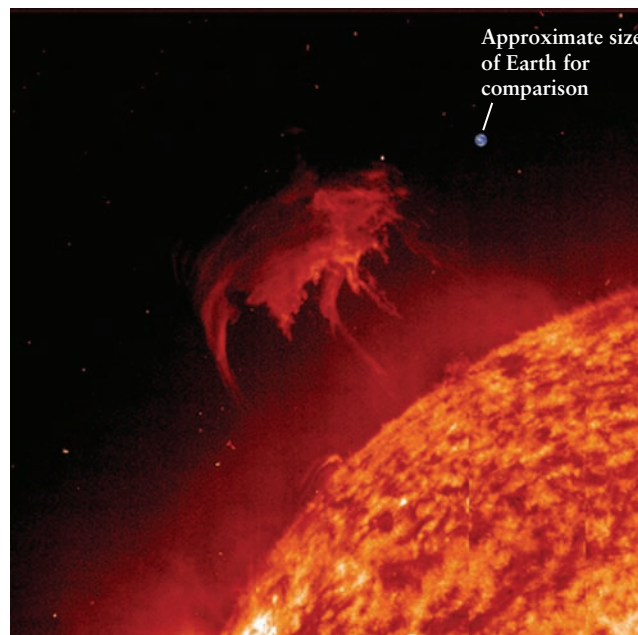
FIGURE 9-14 Active Sun in H_α This photograph shows the chromosphere and lower corona during a solar maximum, when sunspots are abundant. The image was taken through a filter that allowed only light from H_α emission to pass through. The hot, upper layers of the Sun's atmosphere are strong emitters of H_α photons. (See Section 3-17 for details of the Balmer series, $H_\alpha - H_{\gamma}$.) A few large sunspots are evident. Most notable are features that do not appear at the solar minimum, such as the snakelike features shown here called filaments, bright areas called plages, and prominences (filaments seen edge-on) observed at the solar limb. (NASA)



a

R I V U X G

FIGURE 9-15 Prominences (a) A huge prominence arches above the solar surface in this SOHO image taken in 2001. The radiation that exposed this picture is from singly ionized helium at a wavelength of 30.4 nm, corresponding to a temperature of about 50,000 K. (b) The gas in these prominences



b

R I V U X G

was so energetic that it broke free from the magnetic fields that shape and confine it. This coronal mass ejection occurred in 1999 (and did not strike Earth, which is shown for size). (a and b: SOHO [ESA & NASA])

they are hotter, and therefore brighter, than the surrounding chromosphere. Plages, which often appear just before nearby sunspots form, are thought to be created by the magnetic fields under the photosphere crowding upward just before they emerge through the photosphere. In pushing upward, the fields compress the gases of the upper Sun, causing this gas to become hotter and therefore to glow more brightly.

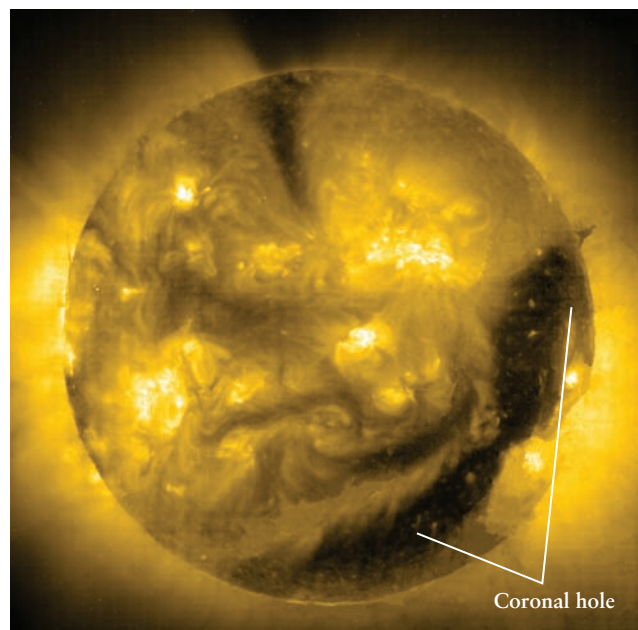
The dark streaks in Figure 9-14 are features in the corona called **filaments**. These features are huge volumes of gas lofted upward from the photosphere by the Sun's magnetic field. When viewed from the side rather than from above, filaments form gigantic loops or arches, called **prominences** (Figure 9-15; see also Figure 9-14 and the figure that opens this chapter). The temperature of the gas in prominences can reach 50,000 K. These features are almost always associated with sunspots. Some prominences last for only a few hours, whereas others persist for months.

Focus Question 9-5

What holds the gas in loops above the Sun's surface?

X-ray photographs also reveal numerous coronal bright and dark spots that are hotter or colder, respectively, than the surrounding corona (Figure 9-16). Temperatures in the bright regions occasionally reach 4 million K. Many of the bright coronal hot spots are located over sunspots. The darker, cooler **coronal holes** act

as conduits for gases to flow out of the Sun. Therefore, when a coronal hole on the rotating Sun faces Earth, the solar wind in our direction increases dramatically.



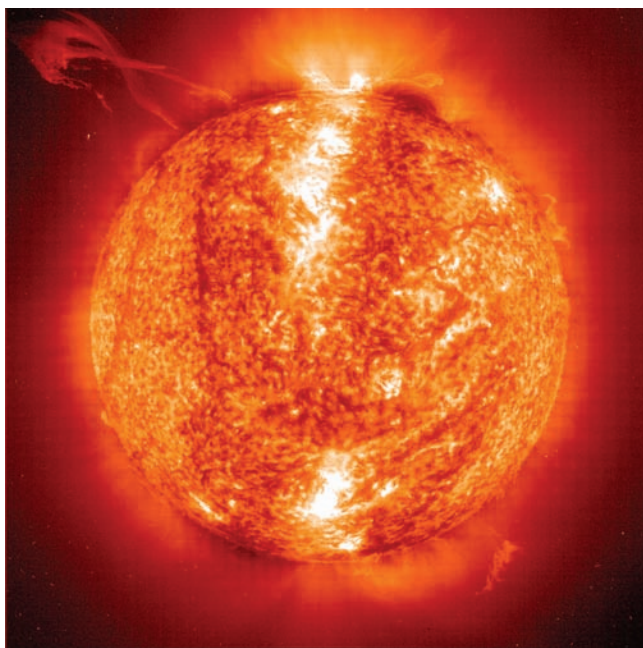
Coronal hole

R I V U X G

FIGURE 9-16 A Coronal Hole This X-ray picture of the Sun's corona was taken by the SOHO satellite. A huge coronal hole dominates the lower right side of the corona. The bright regions are emissions from sunspot groups. (SOHO/ESA/NASA)

Focus Question 9-6

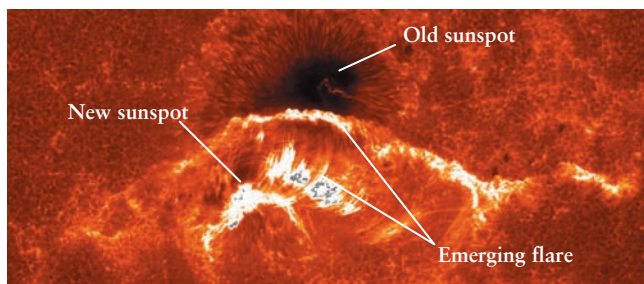
What two features on the Sun are actually the same thing, seen from different angles?



R I V U X G

FIGURE 9-17 A Flare Solar flares, which are associated with sunspot groups, produce energetic emissions of particles from the Sun. This image, taken in 2000 by SOHO, shows a twisted flare (upper left part of figure) in which the Sun's magnetic field lines are still threaded through the region of emerging particles. (NASA)

Violent, eruptive events on the Sun were first discovered to occur in 1859. These **solar flares** release vast quantities of high-energy particles, as well as X rays and ultraviolet radiation (**Figure 9-17**), and they last between a few seconds and 5 hours. Solar flares sometimes occur when sunspots collide (**Figure 9-18**). Typical flares each emit as much energy as is contained in all the fossil fuel ever stored inside Earth. Flares are so powerful that they leave the region of the surface of the Sun in their vicinity quaking for an hour or more. At the maximum of the sunspot cycle, about 1100 flares occur per year. Most of them last for less than an hour, but during that time, temperatures in the Sun's atmosphere soar to 5 million K. By following the paths of particles emitted by flares as



R I V U X G

FIGURE 9-18 Origin of a Solar Flare This *Hinode* satellite image of the Sun shows a developing sunspot colliding with a pre-existing sunspot. The interacting magnetic fields funnel hot gases rapidly away from the Sun as a solar flare. (*Hinode* JAXA/NASA)

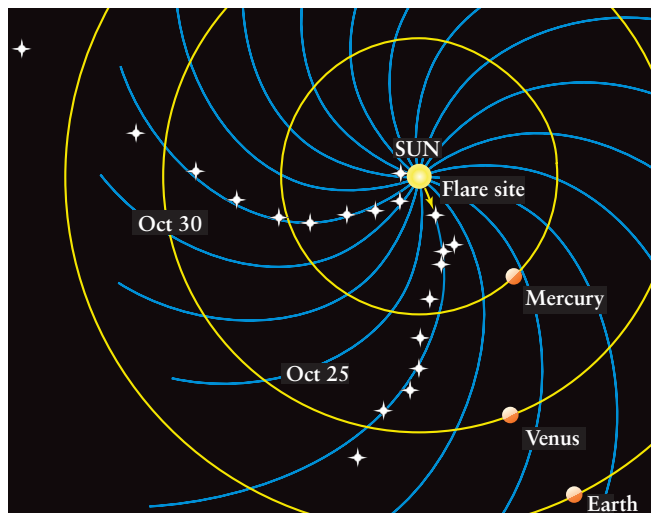


FIGURE 9-19 A Snapshot of the Sun's Global Magnetic Field By following the paths of particles emitted by a solar flare, astronomers have begun mapping the solar magnetic field outside the Sun. The field guides the outflowing particles, which, in turn, emit radio waves that indicate the position of the field. These data were collected by the *Ulysses* spacecraft in 1994. *Ulysses* was the first spacecraft to explore interplanetary space from high above the plane of the ecliptic.

they are deflected by the Sun's magnetic field, astronomers have, since the 1960s, been mapping that field as it extends out beyond Earth (**Figure 9-19**).

Astronomers have also observed huge loop, canopy, or balloon-shaped volumes of high-energy gas being ejected from the corona. These **coronal mass ejections** expel up to 9×10^{12} kg (10 billion tons) of matter at 1.4×10^6 km/hr (9.1×10^5 mi/hr), and each one lasts for up to a few hours (**Figure 9-20**, see also **Figure 9-15b**). Coronal mass ejections have enough energy to break through the Sun's magnetic fields that normally contain them or even to carry fields outward. Flares have been observed to create shock waves that initiate some of the coronal mass ejections. The origins of other ejections are still under investigation. The numbers of plagues, prominences, flares, and coronal mass ejections vary with the same 11-year cycle as the number of sunspots. Coronal mass ejections, the major source of hazardous particles from the Sun, occur with varying frequency throughout the sunspot cycle, but they never completely cease.

Some coronal mass ejections, solar flares, and filaments head toward Earth. Their electromagnetic radiation gets here in about 8 minutes; their high-energy particles arrive a few days later. At times when these surges of particles are *not* coming to us, the normal solar wind particles are trapped by Earth's magnetic fields in the regions called the Van Allen belts. The additional particles from a coronal mass ejection or other solar event that *do* come our way overwhelm the Van Allen belts, enabling matter in them to cascade Earthward. One of the most spectacular of such events in recent years occurred on and around October 31, 2003 (Halloween in

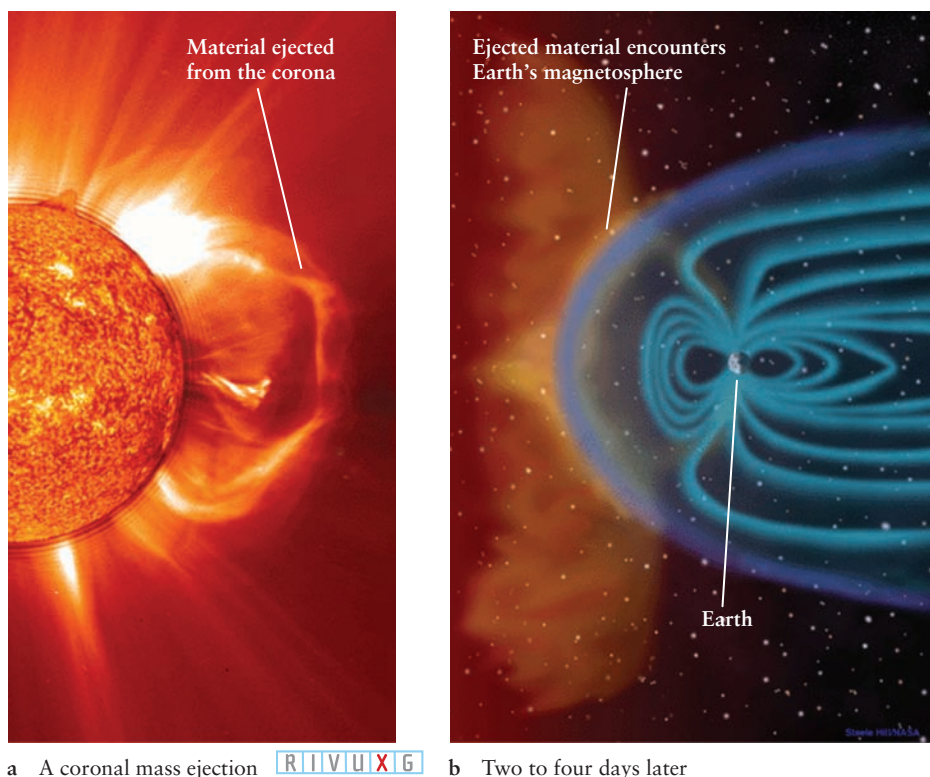


FIGURE 9-20 A Coronal Mass Ejection (a) An X-ray image of a coronal mass ejection from the Sun taken by SOHO. (b) Two to 4 days later, the highest-energy gases from the ejection reach 1 AU. If they come our way, most particles are deflected by Earth's magnetic field (blue). However, as shown, some particles leak Earthward, causing aurorae, disrupting radio communications and electric power transmission, damaging satellites, and ejecting some of Earth's atmosphere into interplanetary space. (a and b: SOHO/ESA/NASA)

the United States), filling the night skies with aurorae in many places around the world (see Section 6-5 for more details on Earth's magnetic fields and aurorae).

Strong surges of high-energy particles from the Sun damage satellites, disrupt radio communications, short out power grids here, produce intense aurorae in Earth's atmosphere, cause some of Earth's atmosphere to gush into interplanetary space, and potentially harm or kill people in space. Collectively, this activity from the Sun that can affect us is called *space weather*.

Using the technology described in Chapter 3, several spacecraft now monitor the Sun and the space between it and Earth. These spacecraft observe gamma-ray, X-ray, ultraviolet, and visible parts of the Sun's spectrum, and they provide both scientific and space weather information. These satellites also provide data that allow astronomers to accurately forecast several days in advance the arrival of dangerous levels of solar particles. In the past few years, modeling the Sun's surface using helioseismology has become so sophisticated that we can even deduce what is happening on the side of the Sun facing away from Earth. As the Sun rotates, any hazardous activity there will sweep into view. By knowing what is coming even weeks ahead of time, astronauts and spacecraft can be better protected than ever before.

THE SUN'S INTERIOR

During the 1800s, geologists and biologists found convincing evidence that Earth must have existed in more or less its present form for at least hundreds of millions of years. This fact posed severe problems for astrophysicists,

because at that time it seemed impossible to explain how the Sun could continue to shine for so long, radiating immense amounts of energy into space as it does. If the Sun were shining by burning coal or hydrogen gas, it would be ablaze for only 5000 years before consuming all of its fuel.

Everyday experience tells us that the Sun is the source of an enormous amount of energy. We have just seen that observations reveal hot gas, intense magnetic fields, and a variety of continuous and transient features on the Sun's surface. But the energy does not come from the surface gas or the magnetic fields—they have no mechanisms to create it. The origin of the Sun's electromagnetic radiation is deep within it. To understand why the Sun shines, we must understand where its energy originates and how that energy is transported to its surface.

9-7 Thermonuclear reactions in the core of the Sun produce its energy

In 1905, Albert Einstein provided an important clue to the source of the Sun's energy with his theory of special relativity. One of the implications of this theory is that matter and energy are related by the simple equation:

$$E = mc^2$$

In other words, a mass (m) can be converted into an amount of energy (E) equivalent to mc^2 , where c is the speed of light. Because c^2 (that is, $c \times c$) is huge, namely, $9 \times 10^{10} \text{ km}^2/\text{s}^2$ ($5.6 \times 10^{10} \text{ mi}^2/\text{s}^2$), a small amount of matter can be converted into an awesome amount of energy.

5 Inspired by Einstein's work, physicists discovered that the Sun's energy output comes from the conversion

of matter into energy. In the 1920s, the British astrophysicist Arthur Eddington proposed that the temperature at the center of the Sun, its **core**, is much greater than had ever been imagined. Calculations eventually revealed that the temperature there is about 15.5×10^6 K. Physicists also showed that at temperatures above about 10×10^6 K, hydrogen fuses into the element helium. In each such reaction, a tiny amount of mass is converted into energy in the form of gamma rays. Enough fusion occurs in its core to account for all of the energy emitted by the Sun.

The process of fusing nuclei at such extreme temperatures is called **thermonuclear fusion** (see **Discovery 9-1: Thermonuclear Fusion**). In particular, conversion of hydrogen into helium is called **hydrogen fusion**. The same process provides the devastating energy released by a hydrogen bomb. (We will encounter the thermonuclear fusion of helium and other elements in later chapters.)

Hydrogen fusion is also called *hydrogen burning*, even though nothing is burned in the conventional sense. The ordinary burning of wood, coal, or any flammable substance is a chemical process involving only the electrons orbiting the nuclei of the atoms. Thermonuclear fusion is a far more energetic process that involves violent collisions that change the atomic nuclei themselves. Discovery 9-1 shows you just how energetic.

6 You may have heard statements like “matter is always conserved” or “energy is always conserved.” We know that both of these concepts are incorrect, because mass can be converted into energy, and vice versa. What is true, however, is that the total amount of energy plus mass (multiplied by c^2) is conserved. So, the destruction of mass and the creation of energy by the Sun do not violate any laws of nature.

Now let us add up the energy emitted from the entire Sun. As noted earlier in this chapter, the Sun’s mass, usually designated $1 M_{\odot}$, is equal to 333,000 Earth masses. Its total energy output per second, called the **solar luminosity** and denoted $1 L_{\odot}$, is 3.9×10^{26} watts. To produce this luminosity, the Sun converts 600 million metric tons of hydrogen into helium within its core each second. This value is twice the mass of the Empire State Building in New York City. This enormous rate is possible because the Sun contains a vast supply of hydrogen—enough to continue the present rate of energy output for another 5 billion years.

9-8 The solar model describes how energy escapes from the Sun’s core

A scientific theory describing the Sun’s interior, called the **solar model**, explains how the energy from nuclear fusion in the Sun’s core gets to its photosphere. The model begins with the inward force due to the Sun’s gravity. This force increases the pressure and temperature in the Sun’s core, thereby causing hydrogen fusion to occur there. Because the Sun is not shrinking today, however, there must be an outward force throughout its interior that counters

the inward force of gravity. That outward force is created by the gamma-ray photons generated during fusion.

Consider an idealized journey of one gamma-ray photon created by fusion in the Sun’s core. The photon begins its journey outward, quickly slamming into a nearby ion. The photon then disappears, its energy being incorporated into the ion, which as a result moves upward at very high speed. Because the particles in the Sun are densely packed, the energized ion does not travel far before striking another particle above it. After the collision, the two particles involved rebound (think billiard balls here). As a result of the collision, a photon is emitted from the interacting particles. The emitted photon has less energy than the photon that was previously absorbed by the ion. The energy taken from the incoming photon in the collision goes into preventing the upper particle, into which the ion slammed, from descending deeper into the Sun. The lower-energy photon slams into another particle above it and the process repeats. For each photon created by fusion, this happens numerous times. Combining these collisions with bulk upward flow of gases (by convection, discussed shortly), the photon energy that did not go into preventing the Sun from collapsing finally gets to the photosphere and is radiated into space as part of sunlight. From fusion to sunlight typically takes about 170,000 years.

Over 10^{38} fusion events occur every second in the Sun’s core, providing just the right amount of energy to prevent it from collapsing. The balance between the inward force of gravity and the outward force from the motion of the hot gas is called **hydrostatic equilibrium** (**Figure 9-21**).

The outward movement of energy by photons hitting particles, which then bounce off other particles and thereby reemit photons, is called *radiative transport*, because individual photons are responsible for carrying energy from collision to collision. Calculations show that radiative transport is the dominant means of outward energy flow in the **radiative zone**, extending from the core to about 70% of the way out to the photosphere.

Near the photosphere, energy is carried the remaining distance to the surface by the bulk motion of the hot gas, rather than by the flying, energetic photons. As we discussed in Section 7-1, this circulation of gas blobs is called *convection*. We thus say that the Sun has a **convective zone**. Hot gas travels up to the top of this zone by convection and from there it radiates photons into space. These photons are what we see as sunlight, and you can see that the photosphere, which is at the bottom of the solar atmosphere, is also at the top of the convective zone. Once the gases in the convective zone lose energy by emitting photons into space, they cool and settle back into the Sun. As noted earlier, this convective flow is also the origin of solar granules, and the departing energy is the visible light and other electromagnetic radiation that the Sun emits into space. **Figure 9-22a** is a sketch of our model of the internal structure of the Sun.

Focus Question 9-7

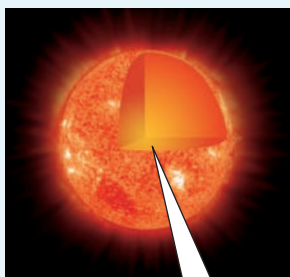
Where in the Sun is the energy that we see generated?

DISCOVERY 9-1

Thermonuclear Fusion

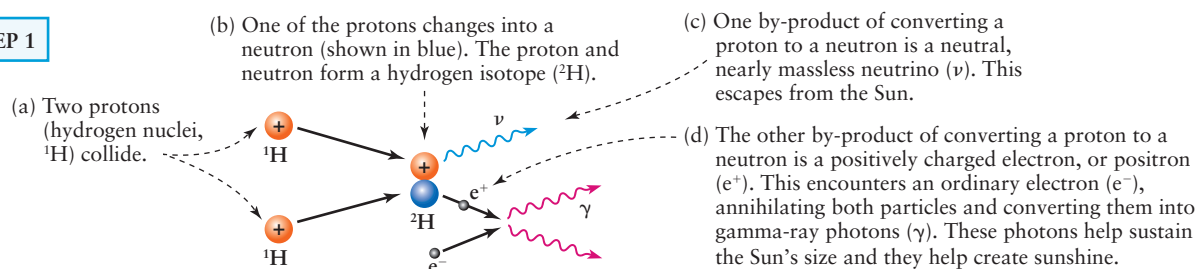
What drives the thermonuclear fusion that powers the Sun? For nuclei to fuse, they must be brought together at incredibly high temperatures and pressures.

That process is exactly what occurs in the Sun's core, where the entire mass of the Sun compresses inward. The core's temperature is 15.5 million K, its pressure

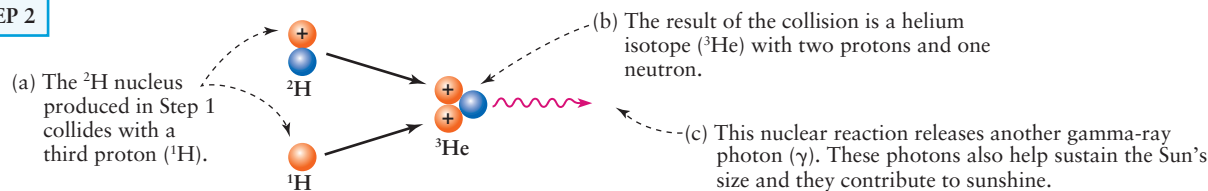


Hydrogen fusion in the Sun usually takes place in a sequence of steps called the proton-proton chain. Each of these steps releases energy that heats the Sun and gives it its luminosity.

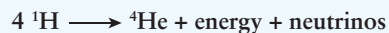
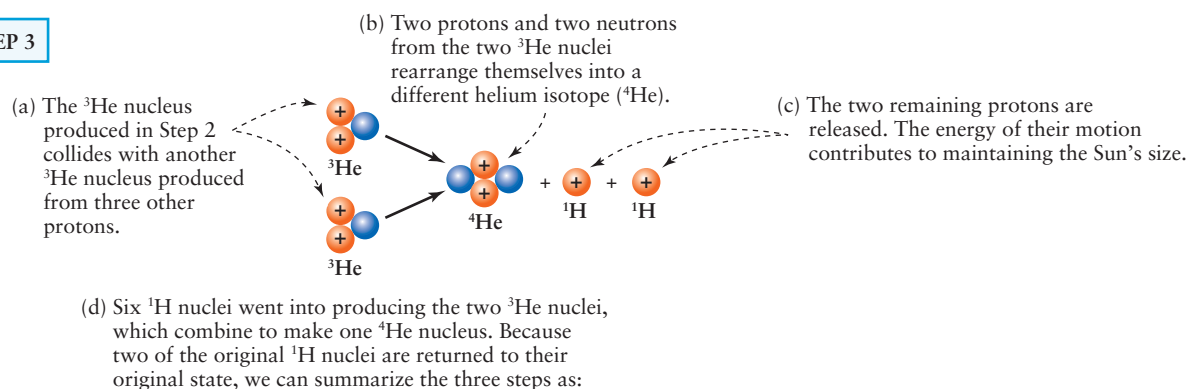
STEP 1



STEP 2



STEP 3



is about 3.4×10^{11} atm, and its density is 160 times greater than that of water.

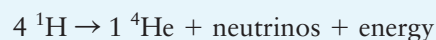
Normally, nuclei cannot contact one another because the positive electric charge on each proton prevents nearby protons from coming too closely together. (Remember that like charges repel each other.) But in the extreme heat and pressure of the Sun's center, the protons move so fast in such close proximity that they can stick, or fuse, together.

The nuclear transformations inside the Sun follow several routes, but each begins with the simplest atom, hydrogen (H). Most hydrogen nuclei consist of a single proton. The final outcome of fusion is the creation of a nucleus of the next simplest atom, helium (He), consisting of two protons and two neutrons. The fusion of hydrogen into helium takes several steps.

The accompanying diagram on the preceding page shows the most common path for hydrogen fusion in the Sun. This particular sequence is called the *proton-proton chain* (PP chain).

Note that proton-proton fusion releases positively charged electrons, e^+ , called **positrons**, and neutral, nearly massless particles, called **neutrinos**, ν . When these positrons encounter regular electrons in the Sun's core, both particles are annihilated, and their mass is converted into energy in the form of gamma-ray photons. The final fusion, in which the helium forms, also returns two protons to the core, which are then available to fuse again.

Because the PP chain produces both neutrinos and the Sun's energy, we can summarize hydrogen fusion this way:



Example: From our summary equation, we can calculate the energy released during a fusion reaction. We simply look at how much mass is converted into energy:

$$\text{Mass of 4 hydrogen atoms} = 6.693 \times 10^{-27} \text{ kg}$$

$$\text{Mass of 1 helium atom} = 6.645 \times 10^{-27} \text{ kg}$$

$$\text{Mass lost} = 0.048 \times 10^{-27} \text{ kg}$$

Thus, a small fraction (0.7%) of the mass of the hydrogen going into the nuclear reactions does not show up in the mass of the helium. Ignoring the extremely small mass of the neutrino, this lost mass is converted into energy, as predicted by Einstein's famous equation:

$$\begin{aligned} E &= mc^2 \\ &= \left(0.048 \times 10^{-27} \text{ kg}\right) \times \left(3 \times 10^8 \text{ m/s}\right)^2 \\ &= 4.3 \times 10^{-12} \text{ J} \end{aligned}$$

Compare! The energy released from the formation of a single helium atom would light a 9-watt lightbulb for almost one-half a trillionth of a second.

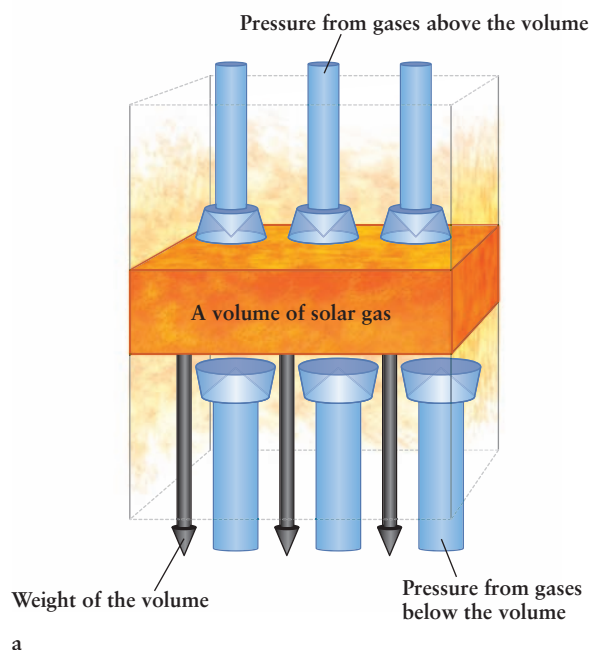


FIGURE 9-21 Hydrostatic Equilibrium (a) Matter deep inside the Sun is in hydrostatic equilibrium, meaning that upward and downward forces (or, equivalently, pressures) on the gases are



balanced. (b) When the pressures on the divers in water are in hydrostatic equilibrium, they neither sink nor rise. (b: Comstock/Stockbyte/Getty Images)

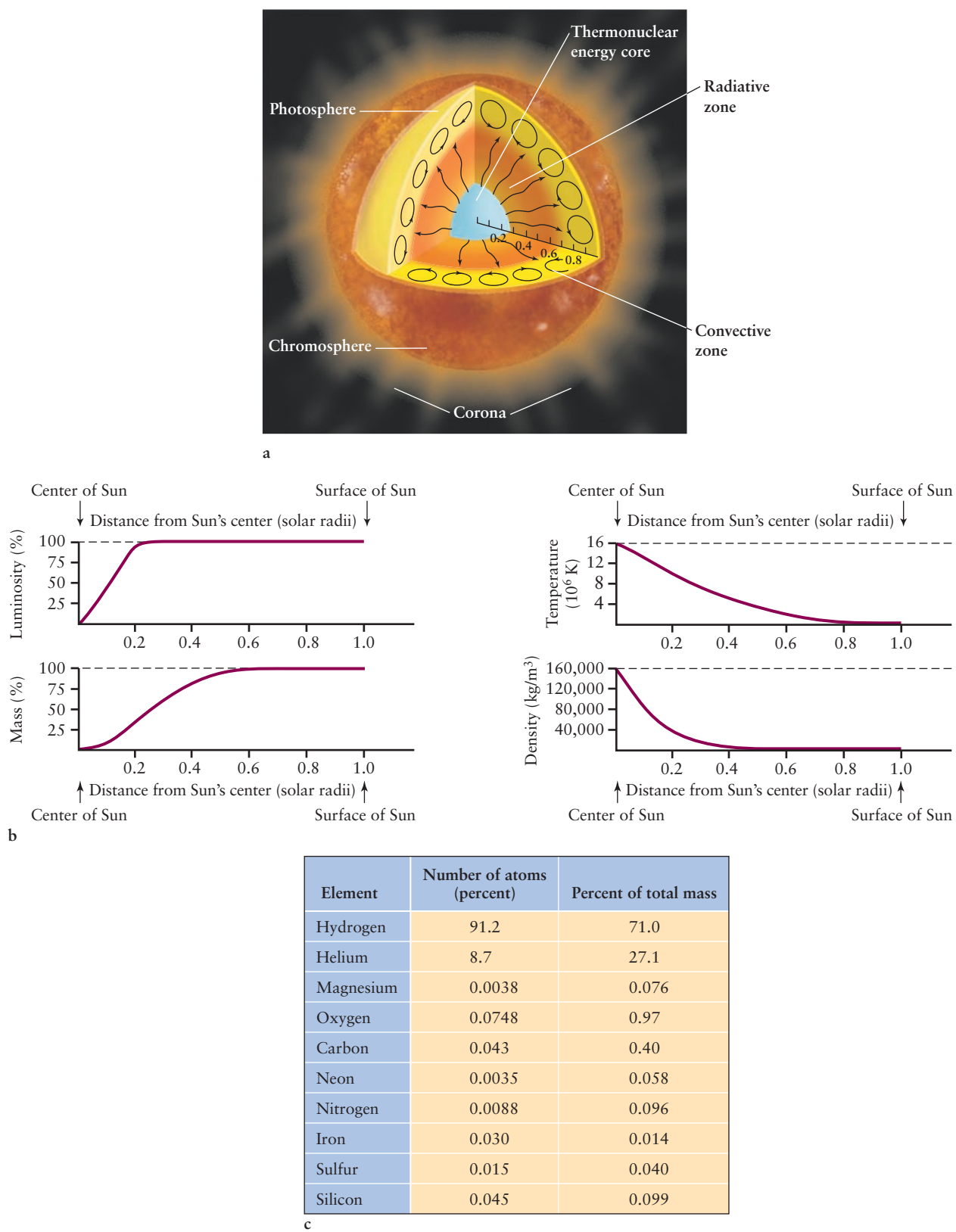


FIGURE 9-22 The Solar Model (a) Thermonuclear reactions occur in the Sun's core, which extends to a distance of 0.25 solar radius from the center. In this model, energy from the core radiates outward to a distance of 0.7 solar radius. Convection is responsible for energy transport in the Sun's outer layers. (b) The Sun's internal structure is displayed here with graphs that show how the luminosity, mass, temperature, and density vary with the distance from the Sun's center. The solar radius (the distance from the Sun's center to the photosphere) equals 696,000 km (432,000 mi). (c) The nine most common elements in the Sun, by the numbers of atoms of each and by the percentage of the Sun's total mass they each comprise. (a: NASA)

Different gamma rays created in the Sun's core lose different amounts of energy as they and their successors travel upward through the Sun. Therefore, the photons emitted from the photosphere have a wide range of energies and, hence, wavelengths. The most intense emission is in the visible part of the electromagnetic spectrum. This process of energy loss by photons traveling up from the Sun's core is the origin of the blackbody nature of the photosphere's spectrum.

Our model of the Sun's interior can be expressed as a set of mathematical equations, called the *equations of stellar structure*. The model quantitatively describes the Sun's internal characteristics, such as its pressure, temperature, and density at various depths. These equations also describe the conditions inside other stars as they evolve. Because the equations are so complex, astrophysicists today use computers to solve them. Figure 9-22b presents their results graphically. The four graphs show how the Sun's luminosity, mass, temperature, and density vary from the Sun's center to its photosphere. For example, the upper left graph gives the percentage of the Sun's luminosity created within that radius. The luminosity increases to 100% at about one-quarter of the way from the Sun's center to its photosphere. This trend tells us that all of the Sun's energy is produced within a volume extending out to $0.25 R_{\odot}$ where $R_{\odot}$ is the radius of the Sun.

The mass curve rises to nearly 100% at about $0.6 R_{\odot}$ from the Sun's center. Almost all of the Sun's mass is therefore confined to a volume extending only 60% of the distance from the Sun's center to its visible surface. This distribution of mass is consistent with the fact that the density of the gas in the photosphere is so low, 10^4 times lower than the density of the air we breathe.

9-9 The Sun has gotten brighter over time

The Sun fuses hydrogen into helium in its core. Therefore, when the Sun first formed, it had less helium and more hydrogen in its core than it does today. Helium is a denser element than hydrogen and therefore the young Sun's core was less dense than it is now. Recall that density is the measure of how much mass is packed into any volume. For example, a cubic meter of carbon (with each atom containing 6 protons and 6 neutrons) is less dense than a cubic meter of iron (with each atom containing 26 protons and 30 neutrons).

When the Sun was young, the larger number of hydrogen atoms (that is, the larger number of protons) in its core created more repulsion than the core experiences today, with its larger number of neutron-rich helium atoms. (Recall that neutrons are neutral and hence do not repel other particles.) The greater repulsion in the young core prevented it from heating up as much as the denser core of today. Therefore, the young Sun, with its cooler core, had less fusion occurring because the cooler a star's core is, the lower the rate of fusion it experiences. Using

the equations of stellar structure on the younger Sun, with its cooler core, reveals that the gamma rays created by fusion back then did not push the outer layers out as much as they do today; hence, the Sun was smaller. But the equations also reveal that the surface of the smaller, younger Sun was about the same temperature as it is today. Because it had less surface from which to radiate, the younger Sun emitted less energy than it does is today—it was dimmer.

Even though its surface temperature has remained roughly constant, the increase in surface area means that the Sun today is giving off more energy than it formerly emitted. In other words, its luminosity has increased. As more helium accumulated in the core, it became denser and was compressed more by the surrounding mass. This increased compression heated the core to higher temperatures, enabling more fusion to occur. The increased numbers of gamma rays thereby generated were able to push outward more and thereby enlarge the Sun to its present size. The Sun has become brighter, meaning that it is emitting more energy than when it was younger. Calculations indicate that the Sun has gotten about 30% more luminous over the past 4.6 billion years.

The fact that the Sun was cooler when the Earth first formed raises the question of whether the Sun always provided enough heat to allow water to be liquid on Earth's surface. There is overwhelming geological evidence for liquid water back at least 4.2 billion years ago. However, if everything on Earth was then as it is today, the young Sun did *not* heat Earth sufficiently for its surface water to be liquid. This is called the *faint young sun paradox*. The paradox is resolved by noting, as we discussed in Section 6-2, that the greenhouse effect created by the Earth's early CO_2 -rich atmosphere kept the Earth sufficiently warm so that surface water was liquid back then.

9-10 The mystery of the missing neutrinos inspired research into the fundamental nature of matter

As explained in Discovery 9-1, for every proton that changes into a neutron during thermonuclear fusion, a *neutrino* is released. Neutrinos have no electric charge, and they are extraordinarily difficult to detect because they rarely interact with ordinary matter. The Sun is largely transparent to neutrinos, allowing these particles to stream outward, unimpeded, from its core. Likewise, most of the neutrinos from the Sun that arrive at Earth pass right through it.

According to our model of fusion in the Sun, nearly 10^{38} *solar neutrinos* per second are produced in the

Focus Question 9-8

Why does the Sun not collapse under the influence of its own enormous gravitational attraction?

Focus Question 9-9

Why did the growing Sun become hotter as measured from Earth if its surface temperature has remained roughly constant?

Sun's core. This output is so huge that here on Earth roughly 100 billion solar neutrinos pass through every square centimeter of your body each second!

Occasionally, however, solar neutrinos strike neutrons and convert them into protons. If astronomers could detect even a few of these converted protons, it might be possible to build a “neutrino telescope” that could be used to detect the thermonuclear inferno in the Sun's core that is hidden from the view of telescopes that collect photons.

Inspired by such possibilities, the American chemist Raymond Davis designed and built a large neutrino detector. This device consisted of a huge tank that contained 100,000 gallons of perchloroethylene (C_2Cl_4 , the fluid your local dry cleaner uses) located deep in the Homestake gold mine in Lead, South Dakota. All neutrino experiments are performed underground to help prevent them from being contaminated by other sources of energy, like high-energy protons from space. Because matter is virtually transparent to neutrinos, most of the solar neutrinos passed right through Davis's tank. On rare occasions, however, a solar neutrino hit the nucleus of one of the chlorine atoms in the cleaning fluid and converted one of its neutrons into a proton, creating a radioactive atom of argon. The rate at which argon was produced was therefore correlated with the number of solar neutrinos arriving at Earth.

On average, solar neutrinos created one radioactive argon atom every 3 days in Davis's tank. To the consternation of physicists and astronomers, this rate corresponds to only one-third of the neutrinos predicted to be created by the fusion in the Sun's core. There were three possible explanations for this unexpected result: The experiment could be faulty; the Sun might not be fusing at the expected rate; or our understanding of the properties of neutrinos could be in error. This experiment, which began in the mid-1960s, was repeated with extreme care by other researchers around the world who got the same results, suggesting that the experiment was not at fault. Calculations of how much energy the Sun must generate to shine as it does confirmed the number of neutrinos created per second. The problem had to lie in our understanding of the neutrino.

Neutrinos, like the planet Neptune, were discovered because, according to theory, they had to exist. Back in the first few decades of the twentieth century, physicists observed neutrons in nuclei spontaneously transforming into protons and electrons. The transformation is a result of the *weak nuclear force*, briefly introduced in Section 3-16. When these scientists calculated the energy and momenta of the neutron and the particles into which it changed, the numbers didn't match—conservation of mass/energy and momentum (see Appendix P: Energy and Momentum) appeared to be violated. If this were indeed the case, it would have required a fundamental reworking of many of the laws

of nature. A simpler solution was proposed in 1930 by the physicist Wolfgang Pauli, who posited the existence of a hard-to-detect particle that carried away the “lost” energy and momentum (this particle was named the *neutrino* in 1931 by physicist Enrico Fermi). Ray Davis shared the Nobel Prize in 2002 for his work in observing solar neutrinos.

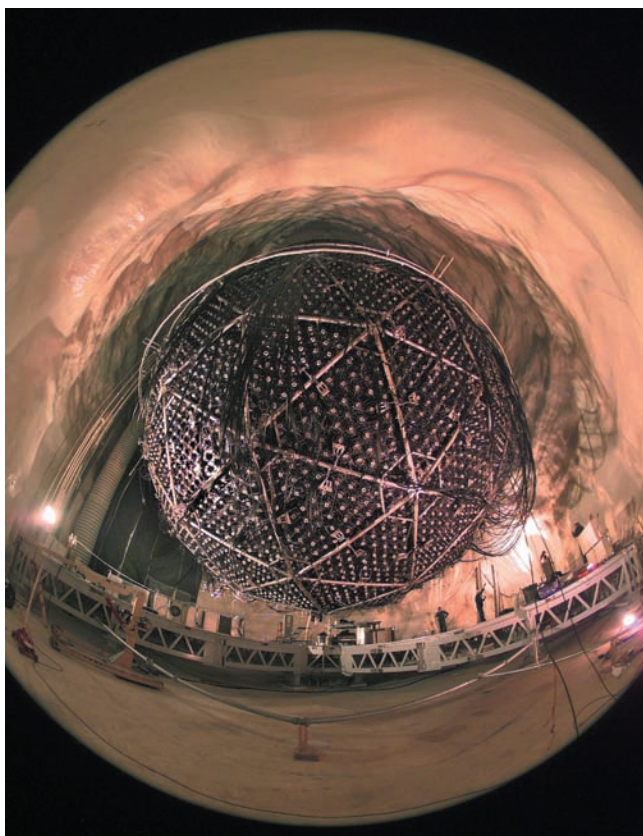
Among the properties neutrinos had to have are very small masses compared to any other types of particles that have mass and very little interaction with other matter, as we have just discussed. Because electrons or positrons (identical to electrons, except with opposite electric charge) always accompany the formation of the neutrinos created in the Sun, they are often called *electron neutrinos*.

Two other kinds of neutrinos were subsequently proposed. One, the *muon neutrino*, is emitted in reactions in which elementary particles, called *muons* (or *antimuons*), are released, while the other, the *tau neutrino*, accompanies the formation of other elementary particles, the tau and antitau particles. The Sun emits neither muon nor tau neutrinos, and the first generation of detectors could only detect electron neutrinos. The solution to the neutrino rate dilemma lay in the existence of muon and tau neutrinos.

Initially, the three “flavors” of neutrinos (electron, muon, and tau) were thought to be massless, and, if so, the equations revealed that they could not change from one flavor into another. But if neutrinos have even the slightest mass, then they can transform from one to another. The transformation of massive neutrinos would occur spontaneously, meaning that if an electron neutrino left the Sun's core and sped toward Earth, there would be a finite probability (about 65%, actually) that by the time it got here, it would have become a tau or muon neutrino. So, if neutrinos have mass, we should observe only one-third of them (which we do), because the rest would have changed flavor before they got to us.

In 1998, the super Kamiokande neutrino detector in Japan observed the other two types of neutrinos coming from the direction of the Sun. The observations were repeated in 2002 at the Sudbury Neutrino Observatory (Figure 9-23a) in Ontario, Canada. At least a dozen other neutrino detectors have begun operations or are being built to further study solar neutrinos.

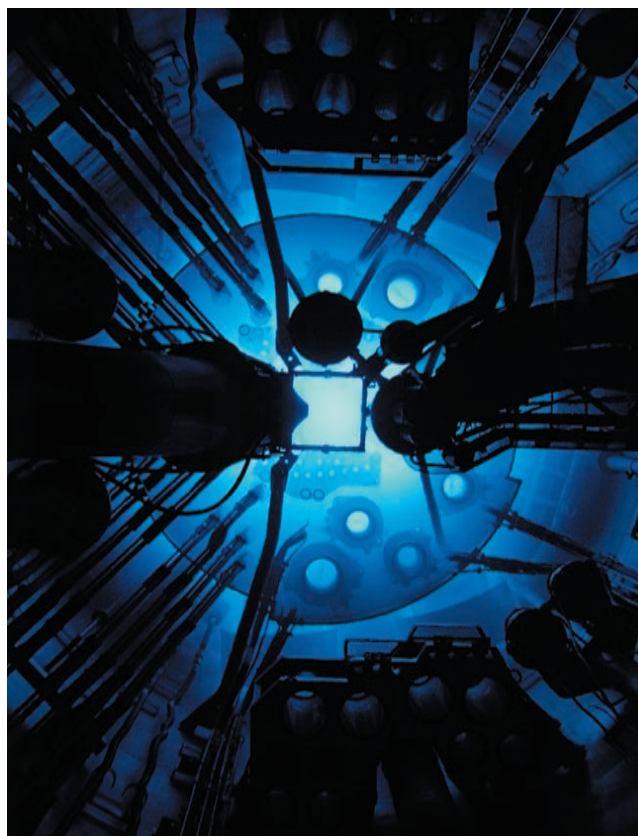
To understand how the present generation of neutrino detectors works, let us consider one of the several interactions that occur between neutrinos and other matter. The Sudbury Neutrino Observatory is filled with water that contains a rare type of hydrogen nucleus, called *deuterium*, which consists of a proton and a neutron. When a solar neutrino is absorbed by a deuterium nucleus, the nucleus breaks apart into two protons and an electron. As this electron rushes through the water, it emits a flash of light, called **Cerenkov radiation** (Figure 9-23b).



a

RIVUXG

FIGURE 9-23 A Solar Neutrino Experiment (a) Located 2073 m (6800 ft) underground in the Creighton nickel mine in Sudbury, Canada, the Sudbury Neutrino Observatory is centered around a tank that contains 1000 tons of water. Occasionally, a neutrino entering the tank interacts with one or another of the particles already there. Such interactions create



b

RIVUXG

flashes of light, called Cerenkov radiation. Some 9600 light detectors sense this light. The numerous silver protrusions are the back sides of the light detectors prior to their being wired and connected to electronics in the lab (seen at the bottom of the photograph). (b) Cerenkov radiation glowing in a nuclear reactor in Australia. (a: LBNL/Science Source; b: ANSTO)

As the Russian physicist Pavel A. Cerenkov (pronounced *Che-REN-kov*) first observed, the flash occurs whenever a particle moves through a medium, such as water, faster than light can. Such motion does not violate the tenet that the speed of light *in a vacuum* (3×10^5 km/s or 186,000 mi/s) is the ultimate speed limit in the universe. Light is slowed considerably as it passes through water, and high-energy particles can exceed this reduced speed of light in a medium without violating the laws of nature.

Thus, scientists detect neutrinos by observing Cerenkov radiation flashes with light-sensitive devices, called *photomultipliers*, mounted in the water (see Figure 9-23a). The three flavors of neutrinos have different interactions with water, all of which have been detected. This evidence that neutrinos can change *requires* that they have mass, and it explains the earlier low rate of solar neutrino observations. The fact that neutrinos have mass was confirmed in 1998, although the amount of mass each has is still under investigation.

SUMMARY OF KEY IDEAS

The Sun's Atmosphere

- The thin shell of the Sun's gases we see are from its photosphere, the lowest level of its atmosphere. The gases in this layer shine nearly as a blackbody. The photosphere's base is at the top of the convective zone.
- Convection of gas from below the photosphere produces features called granules.
- Above the photosphere is a layer of hotter, but less dense, gas called the chromosphere. Jets of gas, called spicules, rise up into the chromosphere along the boundaries of supergranules.
- The outermost layer of gases in the solar atmosphere, called the corona, extends outward to become the solar wind at great distances from the Sun. The gases of the corona are very hot, but they have extremely low densities.

Focus Question 9-10

Why did the early neutrino detectors not detect the predicted number of neutrinos from the Sun?

The Active Sun

- Some surface features on the Sun vary periodically in an 11-year cycle. The magnetic fields that cause these changes actually vary over a 22-year cycle.
- Sunspots are relatively cool regions produced by local concentrations of the Sun's magnetic field protruding through the photosphere. The average number of sunspots and their average latitude vary in the same 11-year cycle.
- A prominence is gas lifted into the Sun's corona by magnetic fields. A solar flare is a brief, but violent, eruption of hot, ionized gases from a sunspot group. Coronal mass ejections send out large quantities of gas from the Sun. Coronal mass ejections and flares that head in Earth's direction affect satellites, communication, and electric power, and cause aurorae.
- The magnetic dynamo model suggests that many transient features of the solar cycle are caused by the effects of differential rotation and convection on the Sun's magnetic field.

The Sun's Interior

- The Sun's energy is produced by the thermonuclear process called hydrogen fusion, in which four hydrogen nuclei release energy when they fuse together to produce a single helium nucleus.
- The energy released in a thermonuclear reaction comes from the conversion of matter into energy, according to Einstein's equation $E = mc^2$.
- The solar model is a theoretical description of the Sun's interior derived from calculations based on the laws of physics. The solar model reveals that hydrogen fusion occurs in a core that extends from the center to about a quarter of the Sun's visible radius.
- Throughout most of the Sun's interior, energy moves outward from the core by radiative diffusion. In the Sun's outer layers, energy is transported to the Sun's surface by convection.
- The amount of energy the Sun emits has increased by about 30% since it first formed.
- Neutrinos were originally believed to be massless. The electron neutrinos generated and emitted by the Sun were originally detected at a lower rate than is predicted by our model of thermonuclear fusion. The discrepancy occurred because electron neutrinos have mass, which causes many of them to change into other forms of neutrinos before they reach Earth. These alternative forms are now being detected.

WHAT DID YOU THINK?

- 1 *What percentage of the solar system's mass is in the Sun?* The sun contains about 99.85% of the solar system's mass.
- 2 *Does the Sun have a solid and liquid interior like Earth?* No. The entire Sun is composed of hot gases.
- 3 *What is the surface of the Sun like?* The Sun has no solid surface. Indeed, it has no solids or liquids anywhere. The level we see, the photosphere, is composed of hot, churning gases.
- 4 *Does the Sun rotate? If so, how fast?* The Sun's surface rotates differentially, varying between once every 35 days near its poles and once every 25 days at its equator.
- 5 *What makes the Sun shine?* Thermonuclear fusion in the Sun's core is the source of the Sun's energy.
- 6 *Are matter and energy conserved?* By themselves, they are not always conserved. Nuclear fusion converts matter into energy. Energy can also be converted into matter. The sum of the matter (multiplied by c^2) and energy is always conserved.

Review Questions

1. Describe the features of the Sun's atmosphere that are always present.
2. Describe the three main layers of the solar atmosphere and how you would best observe them.
3. Name and describe seven features of the active Sun. Which two are the same, seen from different angles?
4. Describe the three main layers of the Sun's interior.
5. Why is the solar cycle said to have a period of 22 years, even though the sunspot cycle is only 11 years long?
6. How do astronomers detect the presence of a magnetic field in hot gases, such as the field in the solar photosphere?
7. Describe the dangers in attempting to observe the Sun. How have astronomers learned to circumvent these hazards?
8. Give an everyday example of hydrostatic equilibrium not presented in the book.
9. Give some everyday examples of heat transfer by convection and radiative transport.

10. What do astronomers mean by a “model of the Sun”?
11. Why do thermonuclear reactions in the Sun take place only in its core?
12. What is hydrogen fusion? This process is sometimes called “hydrogen burning.” How is hydrogen burning fundamentally unlike the burning of a log in a fireplace?
13. Describe the Sun’s interior, including the main physical processes that occur at various levels within the Sun.
14. What is a neutrino, and why are astronomers so interested in detecting neutrinos from the Sun?

Advanced Questions

15. When we are near a sunspot maximum, the Hubble Space Telescope must be moved to a higher orbit. Why? (*Hint:* Think about how the increased solar energy affects Earth’s atmosphere.)
16. Explain how the Sun can be emitting more energy today than shortly after it formed even though its surface temperature has remained roughly constant.

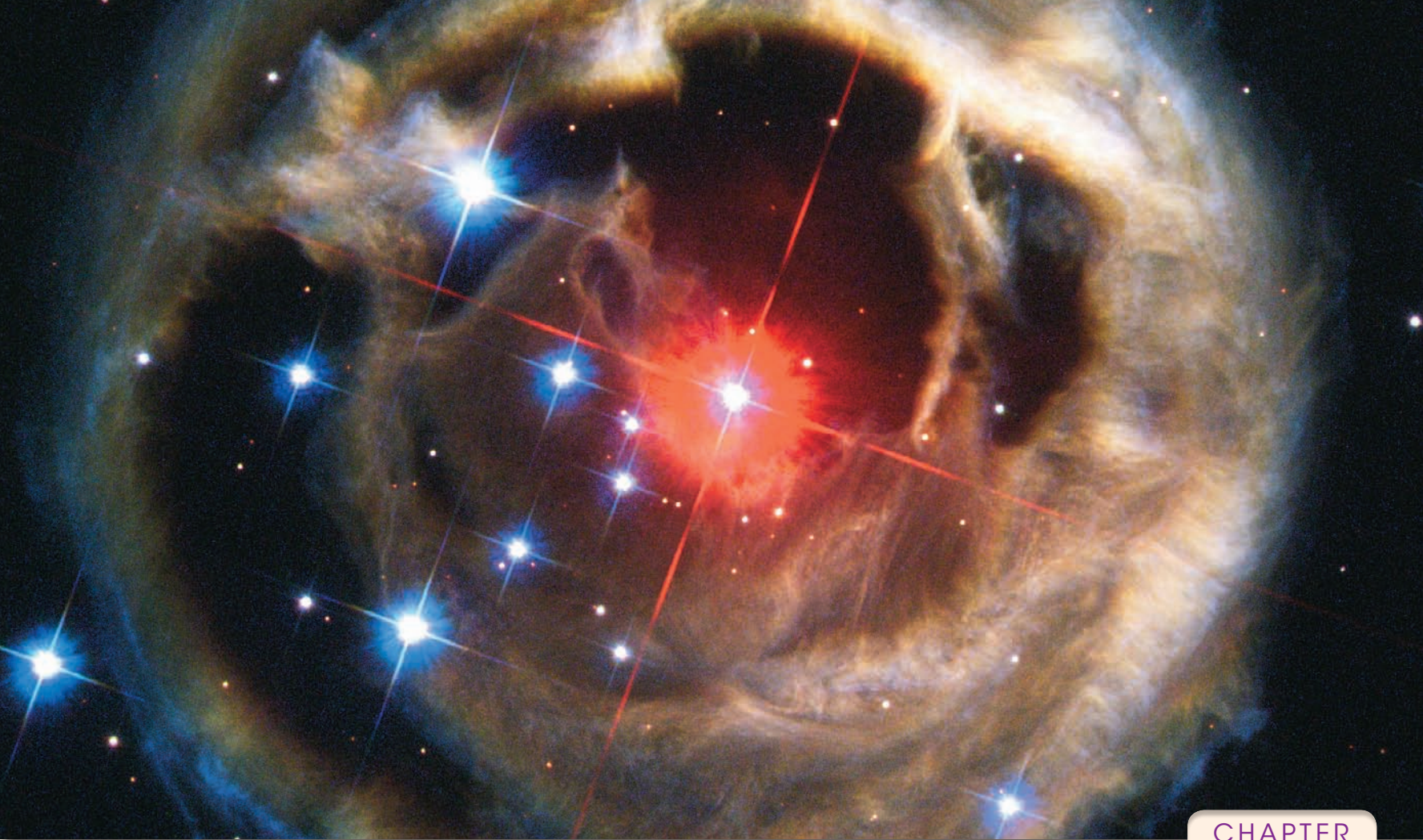
Discussion Questions

17. Discuss the extent to which cultures around the world have worshiped the Sun as a deity throughout history. Why do you think our star inspires such widespread veneration?
18. Discuss some of the difficulties of correlating solar activity with changes in the terrestrial climate.
19. Describe some advantages and disadvantages of observing the Sun (a) from space and (b) from Earth’s South Pole. What kinds of phenomena and issues do solar astronomers want to explore from both Earth-orbiting and Antarctic observatories?

Got It?

20. Does the Sun shine by burning gas, like methane? Justify your answer.
21. What causes sunspots?
22. Does the Sun rotate? Justify your answer.
23. Does the Sun have a solid or liquid surface like Earth?
24. Does the Sun have an atmosphere? Explain your reasoning.

This page intentionally left blank



R I V U X I G Interstellar dust illuminated by light emitted from the red giant star V838 Monocerotis, in the center of the image. (NASA, ESA, and H. E. Bond [STScI])

CHAPTER 10

Characterizing Stars

WHAT DO YOU THINK?

- 1 How near to us is the closest star other than the Sun?
- 2 How luminous is the Sun compared with other stars?
- 3 What colors are stars, and why do they have these colors?
- 4 Are brighter stars hotter than dimmer stars?
- 5 Compared to the Sun, what sizes are other stars?
- 6 Are most stars isolated from other stars, as the Sun is?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

Astronomers are witnesses to literally billions of stars in various stages of evolution. Therefore, although we cannot see any one star go through more than a tiny fraction of the process, we can see stars in every stage of **stellar evolution**. By understanding the changes that stars undergo, we gain insight into our place in the cosmos. We begin our journey to the stars by learning the properties of stars that astronomers used to develop models of why stars shine and how they evolve.

In this chapter you will discover

- that the distances to many nearby stars can be measured directly, whereas the distances to farther ones are determined indirectly
- the observed properties of stars on which astronomers base their models of stellar evolution
- how astronomers analyze starlight to determine a star's temperature and chemical composition
- how the total energy emitted by stars and their surface temperatures are related
- the different classes of stars
- the variety and importance to astronomers of binary star systems
- how astronomers calculate stellar masses

LEAVING THE SOLAR SYSTEM

10-1 Distances to nearby stars are found using stellar parallax

Nothing in human experience prepares us for understanding the distances between the stars. On Earth, the greatest distance we read about in our daily lives is perhaps halfway around the globe, 20,000 km (12,400 mi), which is about the distance from New York City to Perth, Australia. By observing the Sun, the Moon, and planets, we have some comprehension of hundreds of thousands, or even millions, of kilometers or miles.

1 How far away are the nearest stars outside our solar system? Ask 10 people and you will probably hear answers that range from millions to billions of kilometers or miles. In fact, the closest star other than the Sun, Proxima Centauri, in the constellation Centaurus, is about 40 *trillion* km (4.0×10^{13} km or 25 trillion mi) away. Light takes more than 4 years to get here from there. Most of the stars you see in the night sky are many times farther away. Given such enormous separations between astronomical bodies, how do we measure the distance to a nearby star?

Because Proxima Centauri is closer to us than other stars, its position among the background stars changes as Earth orbits the Sun. We see precisely the same effect when, for example, Mars appears to have retrograde motion as we pass between it and the Sun (see Figures 2-2 and 2-3), and it is the effect that Tycho Brahe sought while observing the supernova of 1572 (see Figure 2-7). The apparent motion of nearby stars among the background of more distant stars, due to Earth's orbit around the Sun, is called **stellar parallax**.

Parallax is an everyday phenomenon. We experience it when nearby objects appear to shift their positions against a distant background as we move (review Figure 2-6). We also experience it continuously when we are awake, because our eyes change angle when looking at objects at different distances. As you view a tree 10 m away, your eyes cross only slightly (**Figure 10-1a**). Looking at something closer, say, this book, your eyes cross much more in order for both of them to focus on the same word (**Figure 10-1b**). The parallax angle formed between your eyes and the tree or the book lets your brain judge just how close these objects are to you. Working out distances of nearby stars is done similarly, but it requires painstaking measurements of tiny angles and a little geometry.

As Earth moves from one side of its orbit around the Sun to the other, a nearby star's apparent position shifts among the more distant stars. Referring to Figure 10-1c, the parallax angle, p , is half the angle by which Earth shifts positions through the year as seen from that star, measured in arcseconds. The difference in parallax angles for stars at different distances can be seen by comparing Figures 10-1c and 10-1d. **Discovery 10-1: Distances to Nearby Stars** explores some details of how distances are determined from parallax angles.

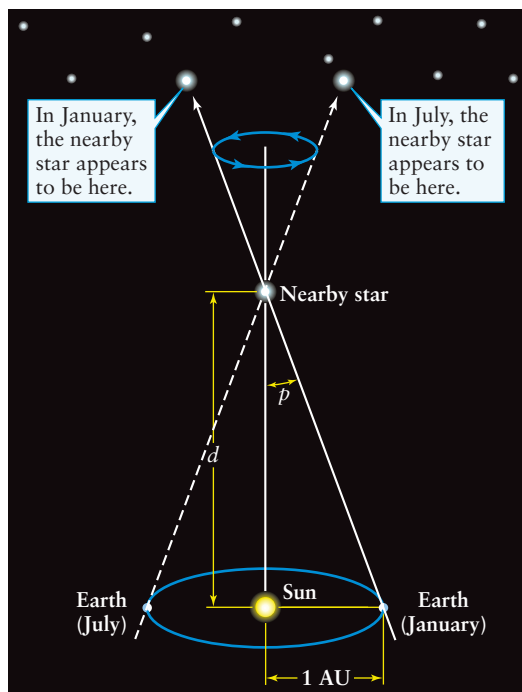
The first stellar parallax measurement was made in 1838 by Friedrich Wilhelm Bessel, a German astronomer and mathematician. He found the parallax angle of the star 61 Cygni to be $\frac{1}{3}''$, and so its distance is about 3 pc (see Discovery 2-2: Units of Astronomical Distance to further review distance units). The precision of stellar parallax measurements is limited by the angular resolution of the telescope, as discussed in Section 3-6. The stellar parallax method using Earth-based telescopes gives stellar distances only up to about 100 pc



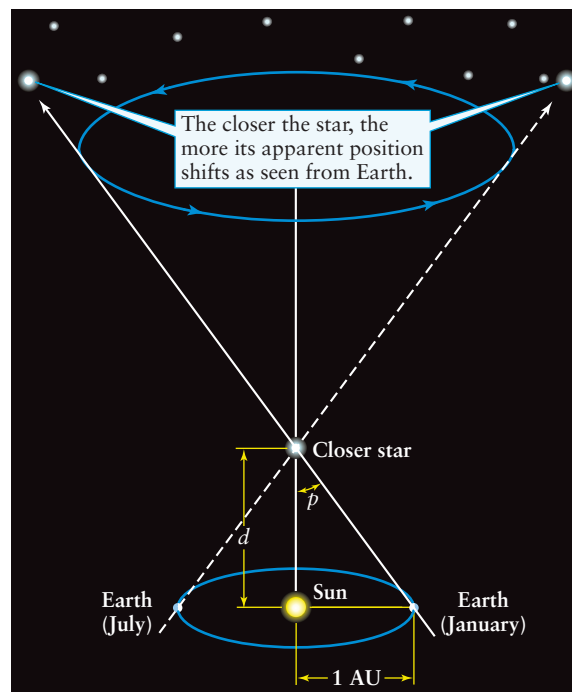
a



b



c Parallax of a nearby star



d Parallax of an even closer star

FIGURE 10-1 Using Parallax to Determine Distance (a, b) Our eyes change the angle between their lines of sight as we look at things that are different distances away. Our eyes are adjusting for the parallax of the things we see. This change helps our brains determine the distances to objects and is analogous to how astronomers determine the distance to objects in space. **(c)** As Earth orbits the Sun, a nearby star appears to shift its position against the background of distant stars. The star's

parallax angle (p) is equal to the angle between the Sun and Earth, as seen from the star. The stars on the scale of this drawing are shown much closer than they are in reality. If drawn to the correct scale, the closest star, other than the Sun, would be about 5 km (3.2 mi) away. **(d)** The closer the star is to us, the greater the parallax angle p . The distance to the star (in parsecs) is found by taking the inverse of the parallax angle p (in arcseconds), $d = 1/p$. (a, b: Mark Andersen/Rubberball/Getty Images)

Focus Question 10-1

For which object—the Moon, Mars, or the star Sirius—is the parallax angle smallest as seen from Earth?

because parallax angles smaller than about 0.01" are difficult to measure from Earth-based observatories.

Telescopes in space are unhampered by our atmosphere and therefore have higher resolutions than Earth-based telescopes. Parallax mea-

surements made in space thus enable astronomers to determine the distances to stars well beyond the reach of ground-based observations.

In 1989, the European Space Agency (ESA) launched a satellite called *Hipparcos* (an acronym for *High Precision Parallax Collecting Satellite* and named for Hipparchus, an astronomer in ancient Greece who created an early classification system for stars). Although the satellite failed to achieve its proper orbit, astronomers used it to measure the distances to more than 2.5 million of the nearest stars up to 500 ly (150 pc) away. The success of *Hipparcos* has led to the 2013 launch of *Gaia* by the European Space Agency. This satellite will use stellar parallax measurements to determine the distances to the nearest 1 billion stars. In 2014, astronomers succeeded in using the parallax method with the Hubble Space Telescope to accurately measure the distances to a few stars 10,000 ly from Earth.

Despite the information gained from stellar parallax, astronomers need to know the distances to more remote stars for which parallax cannot yet be measured. Several methods of determining ever-greater distances will be introduced later in this chapter and in Chapters 11, 12, and 13.

Having established that different stars are at different distances from Earth, we now consider the brightnesses that stars appear to have as seen from our planet. Combining the distances and the varied brightnesses we observe will enable us to calculate how much light stars actually emit and thereby to explore their evolution.

MAGNITUDE SCALES

Greek astronomers, from Hipparchus in the second century B.C.E to Ptolemy in the second century C.E., undertook the classification of stars strictly by evaluating how bright they appear to be relative to each other. This comparison made sense because back then astronomers assumed the stars were all at the same distance from us, and, therefore, differences in brightness due to different stellar distances from Earth were not expected. Classification regardless of distance is still useful to help us navigate around the night sky, and so we will study it

DISCOVERY 10-1

Distances to Nearby Stars

Recall from Discovery 2-2 that 1 parsec (1 pc) is the distance at which two objects 1 AU apart appear 1 arcsec apart. This distance is 3.09×10^{13} km, or 206,265 AU. The word *parsec* (from *parallax second*) originated from the use of parallax to measure distance. Using parsecs, we can write down an especially simple equation for the distances to stars:

Distance to a star in parsecs =

$$\frac{1}{\text{parallax angle of that star in arcseconds}}$$

or

$$d = \frac{1}{p}$$

where d is the distance to the star and p is the parallax angle of that star.

The equation is only this simple in these units, which is one of the main reasons why many astronomers discuss cosmic distances in parsecs rather than light-years. A light-year is the distance light travels through vacuum in one year. We will continue to primarily use light-years

(ly) throughout this book because they are more intuitive. In light-years, this same equation becomes approximately

Distance to a star in light-years $\approx$

$$\frac{3.26}{\text{parallax angle of that star in arcseconds}}$$

or

$$d_{\text{ly}} \approx \frac{3.26}{p}$$

where d_{ly} is the distance to a star in light-years.

Example: The star nearest to the solar system, Proxima Centauri, has a parallax angle of 0.77 arcsec, and so its distance is $1/0.77$, or approximately 1.3 pc. Equivalently, Proxima Centauri is 4.24 ly away. The parallax of Proxima Centauri is comparable to the angular diameter of a dime seen from a distance of 3 km. The parallax angles of the 25 nearest stars are listed in Appendix F: The Nearest Stars.

now. In Section 10-3 we will add the effects of different distances on how bright stars appear from Earth.

10-2 Apparent magnitude measures the brightness of stars as seen from Earth

The brightnesses of stars measured without regard to their distances from Earth are called **apparent magnitudes**, denoted by lowercase m . The brightest stars were originally said to be of first magnitude, and their apparent magnitudes were designated $m = +1$. Those stars that appeared to be about half as bright as first-magnitude stars were said to be second-magnitude stars (designated $m = +2$), and so forth, down to sixth-magnitude stars, the dimmest ones visible to the unaided eye. (Greek astronomers did not try to classify the Sun's dazzling brightness in this scheme.) Because stars do not appear with discrete levels of brightness, this system has noninteger magnitudes as well, such as $+3.5$ or $+4.8$.

More quantitative methods of classifying stars were developed in the mid-nineteenth century. While maintaining the basic idea that the apparent magnitudes of brighter objects are smaller numbers than the apparent magnitudes of dimmer ones, the scale used today begins

by giving the star Vega an apparent magnitude of 0.0. Because some bodies are brighter than Vega, astronomers assign *negative* numbers to the apparent magnitudes of the very *brightest* objects. Sirius, the brightest star in the night sky, has an apparent magnitude of $m = -1.44$. **Figure 10-2a** shows Sirius along with the apparent magnitudes of some of the stars in Orion. With this convention we can describe other bright objects in the sky, such as the Sun, the Moon, comets, and planets. At its brightest, Venus shines with an apparent magnitude of $m = -4.4$, the full Moon has an apparent magnitude of $m = -12.6$, and the Sun has an apparent magnitude of $m = -26.7$. **Remember:** Objects with negative apparent magnitudes appear brighter than those with positive apparent magnitudes—the more negative, the brighter.

Astronomers also extended the magnitude scale so that dimmer stars, visible only through telescopes, could be included in the magnitude system. For example, the dimmest stars visible through a pair of powerful binoculars have an apparent magnitude of about $m = +10$. Time-exposure photographs through telescopes reveal even dimmer stars. The Keck telescopes and the Hubble Space Telescope, among others, image stars as dim as magnitude $m = +30$. Figure 10-2b illustrates the modern

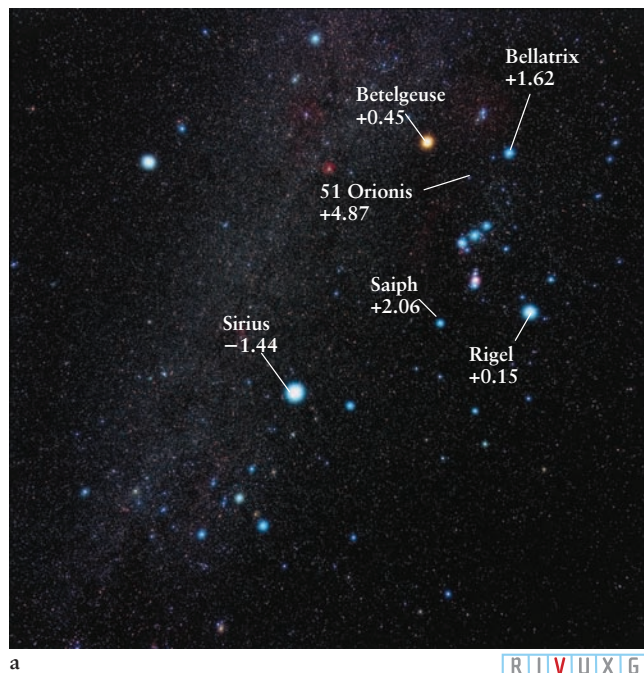
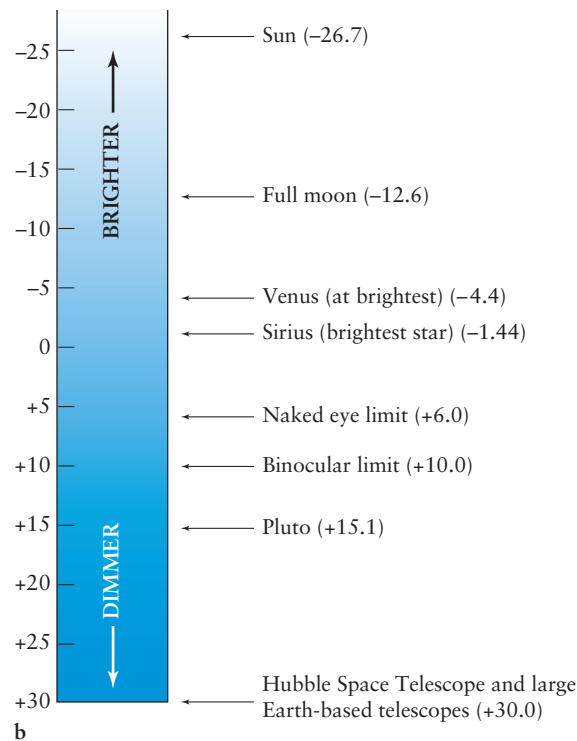


FIGURE 10-2 Apparent Magnitude Scale (a) Several stars in and around the constellation Orion, labeled with their names and apparent magnitudes. (b) Astronomers denote the brightnesses of objects in the sky by their apparent magnitudes. Stars visible to the naked eye have magnitudes between $m = -1.44$ (Sirius) and about $m = +6.0$. However, CCD (charge-coupled device)



photography through the Hubble Space Telescope or a large Earth-based telescope can reveal stars and other objects nearly as faint as magnitude $m = +30$. Note that Pluto's apparent magnitude varies from $+13.7$ to $+16.3$, depending on its distance from Earth. Its average apparent magnitude is $+15.1$. (a: Eckhard Slawik/Science Photo Library/Getty Images)

Insight Into Science

Bigger Is Not Necessarily Brighter Numbering as well as naming schemes may be counterintuitive in science. You might expect brighter stars to have larger, more positive numbers than dimmer stars, but the apparent magnitude scheme is just the opposite. Similarly, we will see shortly that on the standard plot of stars used by astronomers, the Hertzsprung-Russell diagram, the hottest stars fall on the *left* and the coolest stars on the *right*.

apparent magnitude scale. Similarly, apparent magnitudes from entire groups of stars, such as distant galaxies, can be measured.

Knowing, as we now do, that stars are at different distances from us, the apparent magnitudes do not directly reveal fundamental stellar properties. All other things being equal, the closer of two identical stars appears brighter to us (has a smaller apparent magnitude) than the farther star. We take the different distances into account with either of two measures: absolute magnitude and luminosity.

10-3 Absolute magnitudes and luminosities do not depend on distance

To determine the total energy emitted by each star in space, astronomers need to take into account that different stars are at different distances from Earth. They do this by calculating the brightness each star would have if they were all at the same distance from here. Knowing how far away a star really is and how bright it appears (its apparent magnitude), we can calculate how bright it would be at any distance. **Absolute magnitude**, M , is the brightness each star would have at a distance of 10 pc (32.6 ly). Unfortunately, absolute magnitudes have the same counterintuitive numeric scale as apparent magnitudes.

To understand the relationship between the apparent magnitude and the absolute magnitude, we need to know how the brightness of an object changes with distance. Suppose we observe two identical stars, one twice as far away as the other. How much dimmer will the farther one appear to be, as seen from Earth? Light moving

outward from a source spreads over increasingly larger areas of space and, therefore, its brightness decreases with increasing distance. The **inverse-square law** provides the rule for just how quickly the brightness of an object changes with distance.

Imagine light moving out from a star (Figure 10-3a). Start with a small square area of light that has moved out a distance $d = 1$. The light in that square has a certain brightness. When the same square of light has gone

twice as far ($d = 2$), you can see on the figure that it has become 4 times larger. The light in each of the same-sized squares at $d = 2$ contains one-quarter of the photons that were in the single square at $d = 1$. Therefore, each small square at $d = 2$ is one-quarter as bright as the same-sized square at $d = 1$. Similarly, when the square of light has moved to $d = 3$, there are now nine squares the same size as the original, each one-ninth as bright as the original square at $d = 1$. For example, the Sun emits 3.83×10^{26} W (watts) of power (compared to 100 W for a bright incandescent home lightbulb). The Sun's power is spread over a wider area as it travels through space. When it passes Mercury, sunlight provides 9140 watts per square meter (W/m^2) of space. That same energy has spread out so much that by the time it passes us at 1 AU from the Sun, it provides only $1370 \text{ W}/\text{m}^2$. You see the same effect every day when you look at a car's headlights at different distances (Figure 10-3b).

This inverse-square law can be summarized mathematically as follows:

Apparent brightness decreases inversely with the square of the distance between the source and the observer.

Returning to our two identical stars, one twice as far away as the other, the farther one's brightness decreases to $(\frac{1}{2})^2$, or $\frac{1}{4}$, of the brightness of its closer twin by the time the starlight gets to us.

Absolute magnitudes are usually determined by calculations based on apparent magnitudes. Of course, we cannot just move a star to 10 pc distance and then remeasure its apparent magnitude. However, we can calculate the absolute magnitude of a nearby star: We first measure its apparent magnitude and then we find its distance by measuring its parallax angle. Combining these numbers, as shown in **Discovery 10-2: The Distance–Magnitude Relationship**, gives the star's absolute magnitude.

Knowing the Sun's true distance and its apparent magnitude, we can use the inverse-square law to determine how bright it would appear at 10 pc. At that distance, it would have an apparent magnitude of $m = +4.8$. Therefore, the absolute magnitude of the Sun is $M = +4.8$.

2 Because absolute magnitudes tell astronomers how bright stars are compared with one another, this information can be used to evaluate models of stellar evolution, which we discuss in the next few chapters. Absolute magnitudes range from roughly $M = -10$ for the brightest stars to $M = +17$ for the dimmest. Although absolute magnitudes give us comparisons between the energy outputs of stars, we also need to know the total energy they release. The total amount of electromagnetic power (energy emitted each second) is called a star's **luminosity**. The greater the luminosity, the intrinsically brighter the star is compared to stars with lower luminosities. Therefore, the smaller or more negative a

Focus Question 10-2

What might cause the closer of two identical stars to appear dimmer than the farther one?

DISCOVERY 10-2

The Distance–Magnitude Relationship

The closer a star, the brighter it appears. The inverse-square law leads to a simple equation for absolute magnitude, M . Suppose a star's apparent magnitude (corrected for dimming due to interstellar gas and dust between us and the star) is m and its distance from Earth is d (measured in parsecs). Then

$$M = m - 5 \log (d/10)$$

where $\log$ stands for the base-10 logarithm. This distance–magnitude relationship can be rewritten as

$$m - M = 5 \log d - 5$$

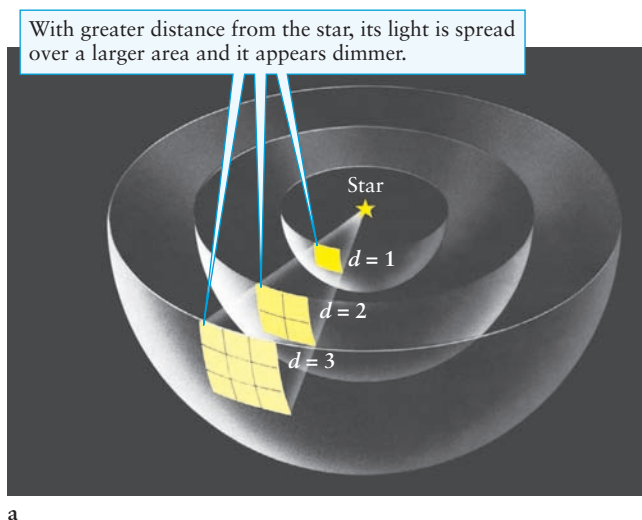
Example: Consider Proxima Centauri, the nearest star to Earth (other than the Sun). By measuring its parallax

angle, we know this star is at a distance from Earth of $d = 1.3$ pc. Its apparent magnitude is $m = +11.1$. Therefore, its absolute magnitude is

$$\begin{aligned} M &= 11.1 - 5 \log (1.3/10) = 11.1 - (-4.4) \\ &= +15.5 \end{aligned}$$

Compare! The Sun is an average star with $M = +4.8$, so Proxima Centauri is an intrinsically dim star. If you know any two of d , m , and M , you can calculate the third variable. For example, if we know a star's absolute and apparent magnitudes, the equation can be used to determine its distance.

star's absolute magnitude, the greater its luminosity. For convenience, stellar luminosities are expressed in multiples of the Sun's luminosity, denoted $L_{\odot}$. As we saw earlier, this value is about 3.83×10^{26} W. The intrinsically brightest stars ($M = -10$) have luminosities of $10^6 L_{\odot}$. In other words, each of these stars has the energy output of a million Suns. The dimmest stars ($M = +17$) have luminosities of $10^{-5} L_{\odot}$. We will provide both luminosity and absolute magnitude data about stars in the chapters that follow.



a

FIGURE 10-3 The Inverse-Square Law (a) The same amount of radiation from a light source must illuminate an ever-increasing area as the distance from the light source increases. The decrease in brightness follows the inverse-square law, which means, for example, that tripling the distance decreases the brightness by a factor of 9. (b) The car is seen at distances of 10 m, 20 m, and 30 m, showing the effect described in part (a). (b: Corbis)

THE TEMPERATURES OF STARS

Armed with knowledge of stellar luminosities, astronomers early in the twentieth century searched for ways to use this information to determine why stars shine. A major step in this process was to find a graph on which physically similar stars are located near each other, because such groupings often provide insight into how objects work. Astronomers found this graph when they plotted either the luminosity or, equivalently, the



30 m away



20 m away



10 m away

b

RTVUXG

Focus Question 10-3
Why can the total energy output of a star be specified equally by its absolute magnitude and its luminosity?

absolute magnitude versus the stars' surface temperatures. The resulting graph led to our models of both stellar activity and stellar evolution. To find the surface temperatures, we start with a fact that is easily overlooked: Stars are not all the same color.

10-4 A star's color reveals its surface temperature

3 Viewing stars with your naked eyes, it is often hard to see that they have different colors, because the color-sensitive cone cells in our eyes are not stimulated very much by such dim objects. Through binoculars or a telescope, their colors, most commonly red, orange, yellow, white, and blue-white, become apparent (see the figure that opens this chapter, Figure 10-2, and **Figure 10-4a**). Like the Sun, all stars are very nearly blackbodies. Therefore, their colors are determined by their surface temperatures. Recall from Figure 3-42 that different blackbody curves are associated with blackbodies of different temperatures. These curves differ in both their intensities and the locations of their peaks, as described by Wien's law (see Section 3-13).

Focus Question 10-4
A star of which temperature has the most intense blue radiation: 3000 K, 5200 K, or 6800 K?

Three typical blackbody curves are presented in Figure 10-4b. The intensity of light from a relatively cool star peaks at long wavelengths, so the star looks red. The intensity of light from a very hot star peaks at shorter wavelengths, making it look blue. The maximum intensity of a star of intermediate temperature, such as the Sun, is found near the middle of the visible spectrum.

To accurately determine the peaks of stars' blackbody spectra and, hence, their surface temperatures, astronomers need to know which blackbody curve most accurately describes each star. **Photometry** provides this information. In this process, a telescope collects starlight that is then passed through one of a set of colored filters and recorded on a charge-coupled device (CCD). At least three photometric images are taken of each star through filters that pass different wavelengths of light. The intensity of a star's image is different at different wavelengths. By plotting the three data points and finding which unique blackbody curve they fit on, we can determine that star's surface temperature.

If a star's surface is very hot, perhaps 10,000 K (see the hottest star in Figure 10-4b), its radiation is skewed toward the ultraviolet, which makes the star bright as detected through an ultraviolet-passing filter, dimmer through a blue filter, and dimmer still through a yellow filter. Regulus, in Leo, is such a star. If a star is cool,

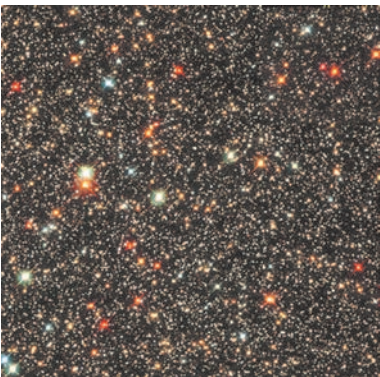
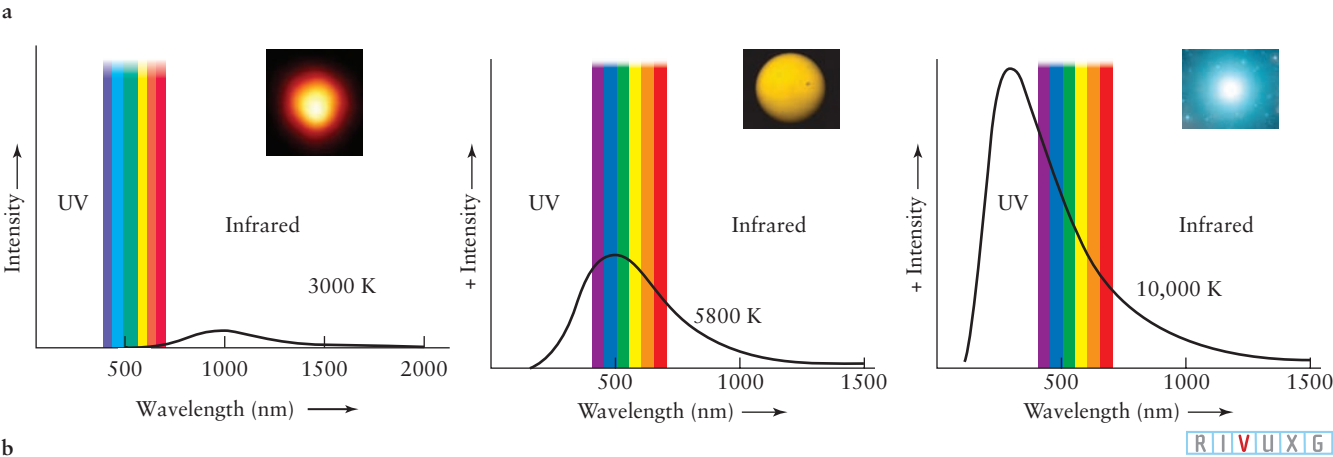


FIGURE 10-4 Temperature and Color (a) This beautiful Hubble Space Telescope image shows the variety of colors of stars. (b) These diagrams show the relationship between the color of a star and its surface temperature. The intensity of light emitted by three stars is plotted against wavelength (compare with Figure 3-42). The range of visible wavelengths is indicated. The location of the peak of each star's intensity curve, relative to the visible-light band, determines the apparent color of its visible light. The insets show stars of about these surface temperatures. Ultraviolet (UV) extends to 10 nm. See Figure 3-7 for more on wavelengths of the spectrum. (a: Hubble Heritage Team(AURA/STScI/NASA); left inset: Andrea Dupree/Harvard-Smithsonian CFA, Ronald Gilliland/STScI, NASA, and ESA; center inset: NSO/AURA/NSF; right inset: Davide De Martin/Science Source)



say, 3000 K (see the coolest star in Figure 10-4b), its radiation peaks at long visible or even infrared wavelengths, making the star bright through a red filter, dimmer through a yellow filter, and dimmer still through a blue filter. Betelgeuse (Figure 10-2a) and Aldebaran are examples of such cool stars. Humans are blackbody radiators with peak temperatures around 300 K. Our peak is in the infrared.

10-5 A star's spectrum also reveals its surface temperature

Another way to determine a star's surface temperature is from a study of its spectrum, a technique called **stellar spectroscopy** (see Chapter 3). Recall that spectral lines result from the absorption and scattering of starlight by gases in the star's atmosphere, in interstellar space, and in Earth's atmosphere. Astronomers begin by taking the spectrum of a star and then identifying and discarding spectral lines due to interstellar gas and Earth's atmosphere. What remain are spectral lines created in the

star's atmosphere. These lines provide us with knowledge of the chemical compositions of the outer layers of stars.

At first glance, there seems to be a bewildering variety of stellar spectra, some of which are shown in **Figure 10-5**. Some stellar spectra show prominent absorption lines of hydrogen. Others exhibit many absorption lines of calcium and iron. Still others are dominated by broad absorption lines created by molecules, such as titanium oxide.

Fortunately, these spectra all hold clues about the nature of stars and, in particular, their surface temperatures. Consider the abundant hydrogen gas found in nearly every star's atmosphere. Although hydrogen accounts for about three-quarters of the mass of a typical star, strong (meaning dark) hydrogen absorption lines do not show up in every star's spectrum. The strength of the absorption lines depends on the star's temperature. In the late 1920s, astronomer Cecilia Payne-Gaposchkin and physicist Meghnad Saha succeeded in explaining why a star's visible hydrogen spectrum is affected by its surface temperature.

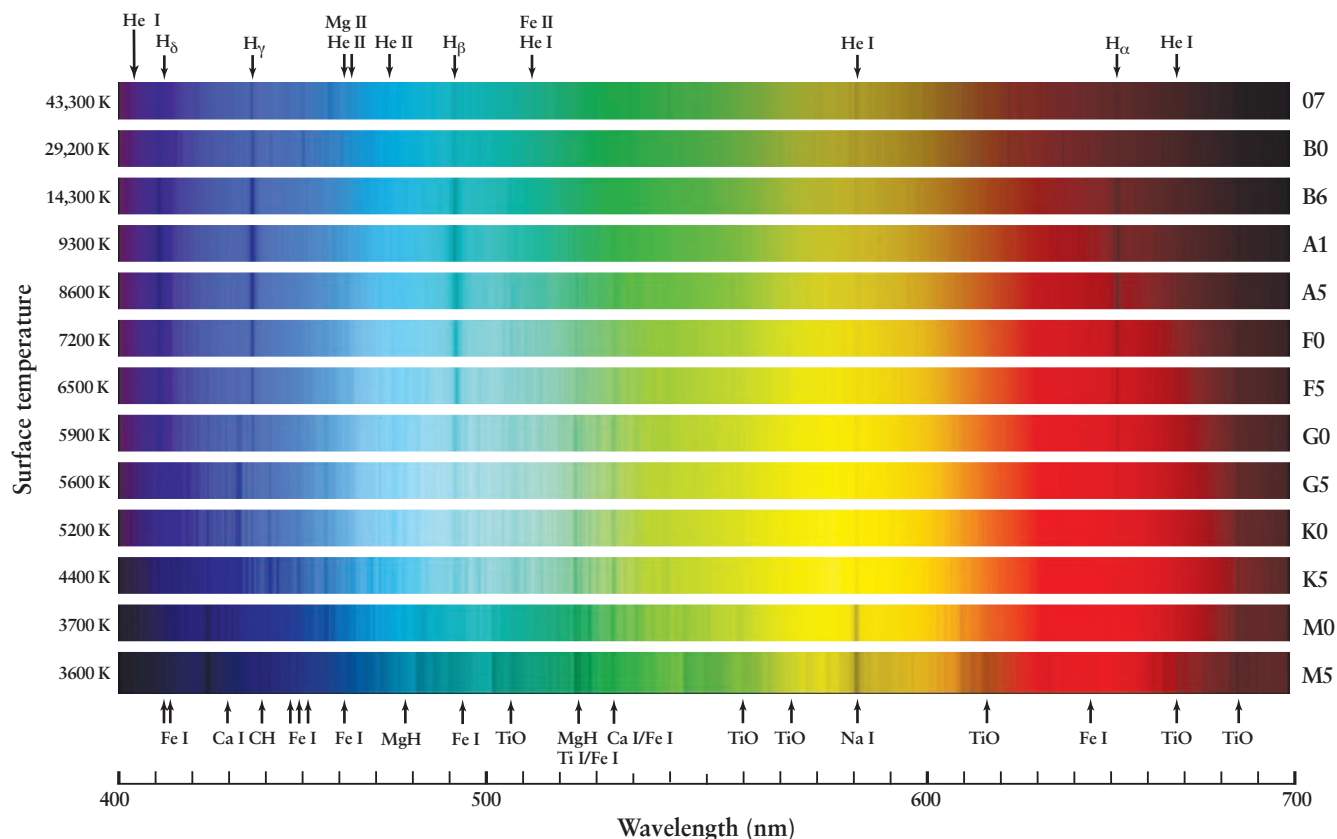


FIGURE 10-5 The Spectra of Stars with Different Surface Temperatures The corresponding spectral types are indicated on the right side of each spectrum. (Note that stars of each spectral type have a range of temperature.) The hydrogen Balmer lines are strongest in stars with surface temperatures of about 10,000 K (called A-type stars). Cooler stars (G- and K-type stars) exhibit numerous atomic lines caused by various elements, indicating temperatures from 4000 to 6000 K.

Several of the broad, dark bands in the spectrum of the coolest stars (M-type stars) are caused by titanium oxide (TiO) molecules, which can exist only if the temperature is below about 3700 K. Recall from Section 3-16 that the Roman numeral I after a chemical symbol means that the absorption line is caused by a neutral atom; a numeral II means that the absorption is caused by atoms that have each lost one electron. (NOAO/Science Source)

Niels Bohr's model of the hydrogen atom (recall Figure 3-52) shows why. As Bohr proposed in the early 1900s, the visible hydrogen lines, called *Balmer lines*, are produced when photons excite electrons initially in the second, $n = 2$, energy level of hydrogen to a higher energy level. Recall from Section 3-17 that these are labeled H_α , H_β , H_γ , and so on. At 10,000 K, most hydrogen electrons absorb enough energy from photons to normally reside in the $n = 2$ level. When these electrons gain further energy, they move to higher energy levels. Therefore, a star with that surface temperature produces the strongest Balmer lines.

Why do hotter or cooler stars produce weaker (less dark) Balmer absorption lines? Suppose first that the star is much hotter than 10,000 K. High-energy photons that stream through the photosphere completely strip away (ionize) electrons from most of the hydrogen atoms there. Because an ionized hydrogen atom cannot produce spectral lines, a very hot star has very dim hydrogen Balmer lines, even though it contains great quantities of hydrogen. Conversely, if a star is much cooler than 10,000 K, most of the photons that pass through its photosphere possess too little energy to keep many electrons there in the $n = 2$ (first excited state) of the hydrogen atoms or to boost electrons in the $n = 2$ state to higher energy states. Therefore, these stars also produce dim Balmer lines.

To produce strong Balmer lines, a star must be hot enough to excite electrons out of the $n = 2$ state but not hot enough to ionize a significant fraction of the atoms.

Focus Question 10-5

Why do very hot stars have weak hydrogen lines, even though they are composed primarily of hydrogen?

As you can see in Figure 10-5, at temperatures much cooler than 10,000 K, the spectral lines of other elements dominate a star's spectrum. For example, the spectral line of neutral sodium, Na I, while dim at around 10,000 K, is quite strong at around 3700 K because photons have enough energy to excite electrons in sodium atoms without ripping them out of orbit. By surveying the relative strengths of a variety of absorption lines, astronomers can determine a star's surface temperature to high precision.

10-6 Stars are classified by their spectra

We have seen that astronomers can determine a star's surface temperature from either the peak of its black-body curve or the strength of its various spectral lines. To make use of the diverse stellar spectra, astronomers since the middle of the nineteenth century have grouped similar spectra into classes, or **spectral types**. According to one classification scheme popular in the late 1800s, a star was assigned a letter between A and P, depending on the strength of the Balmer hydrogen lines in the star's spectrum. It was assumed that the strength of these lines was uniquely related to the star's surface temperature, but as we have seen in the preceding section, this belief is unjustified.

In the early 1900s, Edward C. Pickering and Williamina Fleming, followed by Annie Jump Cannon, and their colleagues at Harvard Observatory (**Figure 10-6**) set up the spectral classification scheme we use today. Many of the early A through P categories were dropped because the Balmer lines in a star's spectrum can be weak



a



b

FIGURE 10-6 Classifying the Spectra of Stars The modern classification scheme for stars, based on their spectra, was developed at the Harvard College Observatory in the late nineteenth century. Female astronomers, initially led by (a) Edward C. Pickering [Back row (L to R): Margaret Harwood (far left), Mollie O'Reilly, Edward C. Pickering, Edith Gill, Annie Jump Cannon, Evelyn Leland (behind Cannon), Florence Cushman, Marion Whyte (behind Cushman), Grace Brooks. Front

row: Arville Walker, unknown (possibly Johanna Mackie), Alta Carpenter, Mabel Gill, Ida Woods] and then by (b) Annie Jump Cannon, analyzed hundreds of thousands of spectra. Social conventions of the time prevented most female astronomers from using research telescopes or receiving salaries comparable to those of men. (a: Emilio Segre Visual Archives/American Institute of Physics/Science Photo Library/Science Source; b: © Bettmann/Corbis)

whether the star is very cool or very hot. The remaining Balmer-based spectral types were thus reordered by stellar surface temperature into the **OBAFGKM sequence**. This sequence is most easily learned with the aid of a mnemonic, such as “Oh, Be A Fine Guy (or Girl), Kiss Me!”

The hottest stars have surface temperatures of 30,000 to 50,000 K and are classified as O type; their spectra are dominated by He II (singly ionized helium) and Si IV (triply ionized silicon). The coolest stars have surface temperatures of 2500 to 3000 K and are classified as M type. **Table 10-1** includes representative examples of each spectral type.

The wide range of temperatures covered by each spectral type prompted astronomers to further subdivide the OBAFGKM temperature sequence. Each spectral type is now broken up into 10 temperature sub-ranges. These 10 finer steps are indicated by adding an integer from 0 (hottest) through 9 (coolest). Thus, an A8 star is hotter than an A9 star, which is hotter than an F0 star, which is hotter than an F1 star, and so on. The Sun, whose spectrum is dominated by singly ionized metals (especially Fe II and Ca II), is a G2 star.

TYPES OF STARS

Keeping in mind that patterns of objects often reveal information about them, early-twentieth-century astronomers began searching for relationships between different properties of stars. The relationship between luminosities and surface temperatures turned out to be a key to understanding types of stars. As often happens in such scientific quests, the significance of this relationship was discovered nearly simultaneously by two independent researchers.

Insight Into Science

Tolerate Idiosyncrasies In astronomy, the word *metal* applies to all elements other than hydrogen or helium. In chemistry, elements are classified as metals or nonmetals according to their physical and chemical properties. To a chemist, sodium and iron are metals, whereas carbon and oxygen are not.

10-7 The Hertzsprung-Russell diagram identifies distinct groups of stars

Around 1911, the Danish astronomer Ejnar Hertzsprung noticed that patterns emerge when the luminosities of stars (or their equivalent absolute magnitudes) are plotted against their surface temperatures (or their equivalent spectral types). Luminosity and absolute magnitude each indicate the total energy emitted by a star or other body. These two terms provide quantitative measurements in answer to the question: “How bright is that star?” Within two years, the American astronomer Henry Norris Russell independently discovered the same result. Graphs of stellar luminosity or absolute magnitude against surface temperature or spectral type are now known as **Hertzsprung-Russell diagrams**, or **H-R diagrams**.

The H-R diagram is valuable because it shows that stars do not have random surface temperatures and luminosities; the two factors are correlated. **Figure 10-7** is a typical H-R diagram. Each dot represents a star whose luminosity and spectral type have been determined. The surface temperatures are plotted along the top of this

Focus Question 10-6

Which star is hottest: F5, B6, or M3?

TABLE 10-1 THE SPECTRAL SEQUENCE

Spectral class	Color	Temperature (K)	Spectral lines	Examples
O	Blue-violet	50,000–30,000	Ionized atoms, especially helium	Naos (ζ Puppis), Mintaka (δ Orionis)
B	Blue-white	30,000–11,000	Neutral helium, some hydrogen	Spica (α Virginis), Rigel (β Orionis)
A	White	11,000–7500	Strong hydrogen, some ionized metals	Sirius (α Canis Majoris), Vega (α Lyrae)
F	Yellow-white	7500–5900	Hydrogen and ionized metals, such as calcium and iron	Canopus (α Carinae), Procyon (α Canis Minoris)
G	Yellow	5900–5200	Both neutral and ionized metals, especially ionized calcium	Sun, Capella (α Aurigae)
K	Orange	5200–3900	Neutral metals	Arcturus (α Boötis), Aldebaran (α Tauri)
M	Red-orange	3900–2500	Strong titanium oxide and some neutral calcium	Antares (α Scorpii), Betelgeuse (α Orionis)

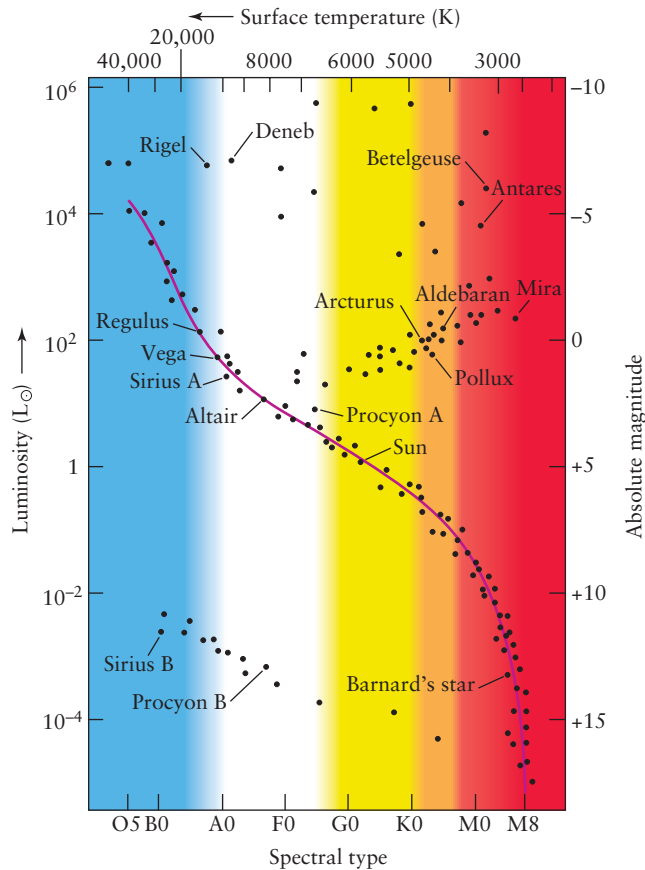


FIGURE 10-7 A Hertzsprung-Russell Diagram On an H-R diagram, the luminosities of stars are plotted against their spectral types. Each dot on this graph represents a star whose luminosity and spectral type have been determined. Some well-known stars are identified. The data points are grouped in just a few regions of the diagram, revealing that luminosity and spectral type are correlated: Main-sequence stars fall along the red curve, giants are to the right, supergiants are on the top, and white dwarfs are below the main sequence. The absolute magnitudes and surface temperatures are listed at the right and top of the graph, respectively. These are sometimes used on H-R diagrams instead of luminosities and spectral types.

figure and the absolute magnitudes along the right side. You can therefore see the equivalence of the temperature and spectral type and the equivalence of the luminosity and absolute magnitude.

Bright stars are near the top of the diagram; dim stars are near the bottom. Contrary to intuition, hot (O and B) stars are toward the left side and cool (M) stars are toward the right. Hertzsprung and Russell made this choice because of the standard sequence OBAFGKM.

The band of stars in Figure 10-7 stretching diagonally across the H-R diagram and on which a red curve is superimposed represents most of the stars we see in the nighttime sky. Called the **main sequence**, it extends from the hot, bright, bluish stars in the upper left corner of the diagram down to the cool dim stars (called *red dwarfs*) in the lower right corner. Each star on this band

Insight Into Science

From Patterns to Models Scientists look at patterns of behavior in related objects as valuable clues to underlying properties and their causes. Guided by these data, they then create mathematical models, which are used to make fresh predictions. For example, astronomers analyze the relationship between the luminosities and surface temperatures of stars as displayed on an H-R diagram to gain insight into the internal activities of stars.

is a **main-sequence star**. Just over 91% of the stars surrounding the solar system fall into this category.

Observations reveal that the number of main-sequence stars decreases with increasing surface temperature. Therefore, along the main sequence, the cooler M, K, and G stars are the most common ones and the hot O stars are the rarest. The Sun (spectral type G2, absolute magnitude +4.8) is a main-sequence star.

4 To the right of the main sequence on the H-R diagram is a second major grouping of stars. These stars are bright but cool. From the Stefan-Boltzmann law (see Section 3-13), we know that a cool object radiates much less light from each unit of surface area than does a hot object. To be so luminous, these cool stars must therefore be huge compared to main-sequence stars of the same temperature, so they are called **giant stars**.

Giants are typically 10 to 100 times the radius of the Sun and have surface temperatures between 2000 and 20,000 K. The cooler members of this class of stars (those with surface temperatures between 2000 and 4500 K) are often called **red giants** because they appear reddish in the nighttime sky. Aldebaran in the constellation Taurus and Arcturus in Boötes are examples of red giants that you can easily see with the naked eye.

A few rare stars are considerably bigger and more luminous than typical giants. Located along the top of the H-R diagram, these superluminous stars are appropriately called **supergiants**. They can extend in radius up to about 1000 $R_{\odot}$. Betelgeuse in Orion and Antares in Scorpius are two examples that are visible in the nighttime sky. Together, giants and supergiants comprise less than 1% of the stars in our vicinity.

5 The remaining 8% of stars in our neighborhood of space fall in a final grouping toward the lower left and bottom of the H-R diagram. As their placement on the diagram shows, these stars are hot, dim, and tiny compared to the Sun. Called **white dwarfs**, we will see that they are actually remnants of stars. White dwarfs are roughly the same size as Earth, and because of their low luminosities, they can be seen only with the aid of a telescope.

The H-R diagram also provides us with enough data to determine a star's radius. There is a relationship between a star's luminosity, L , temperature, T , and

radius, r : $L = \sigma T^4 4\pi r^2$, where σ (sigma) is a constant. Knowing two properties, the luminosity and temperature, we can solve this equation for the third, namely, the radius. Lines of constant stellar radii are drawn on **Figure 10-8**.

These results are summarized in **Figure 10-8**, where the dashed lines indicate the radii of stars. Notice that most main-sequence stars are roughly the same size as the Sun. It is important to keep in mind that, although the H-R diagram is invaluable in helping astronomers *organize* stars by their physical properties (temperature and absolute magnitude), the diagrams do not *explain* the physical mechanisms that produce these characteristics. As we will see shortly, the locations of stars on the H-R diagram provided crucial clues that astrophysicists used in developing and testing their theories of how stars shine.

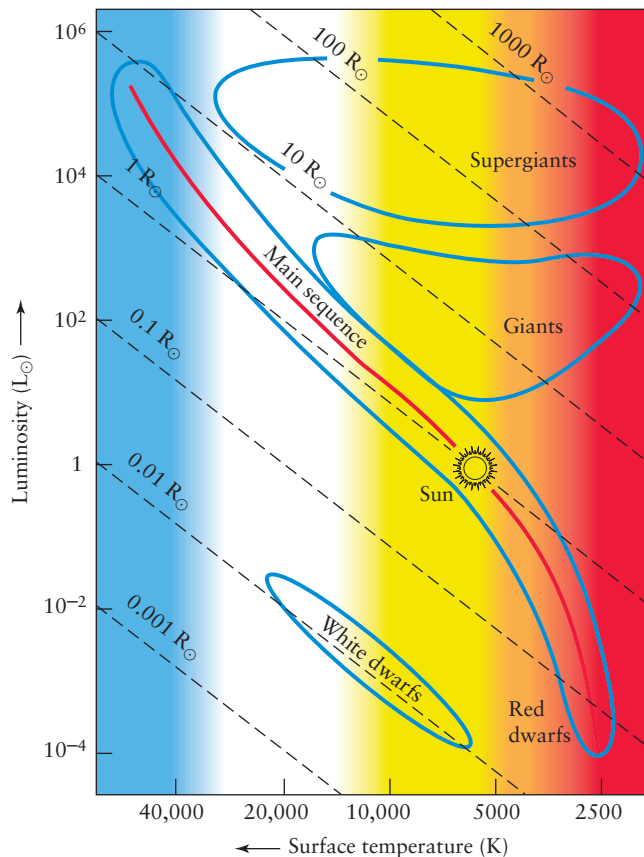


FIGURE 10-8 The Types of Stars and Their Sizes On this H-R diagram, stellar luminosities are plotted against the surface temperatures of stars. The *dashed diagonal lines* indicate stellar radii. For stars of the same radius, hotter stars (corresponding to moving from right to left on the H-R diagram) glow more intensely and are more luminous (corresponding to moving upward on the diagram) than cooler stars. While individual stars are not plotted, we show the regions of the diagram in which main-sequence, giant, supergiant, and white dwarf stars are found. Note that the Sun is intermediate in luminosity, surface temperature, and radius; it is very much a middle-of-the-road star.

10-8 Luminosity classes set the stage for understanding stellar evolution

We have seen that a star's surface temperature largely determines which lines are prominent in its spectrum. Therefore, as noted in Section 10-6, classifying stars by spectral type is the same as categorizing them according to surface temperature. However, as you can see in **Figure 10-8**, stars of the same surface temperature can have different luminosities. For example, a star with a surface temperature of 5800 K could be a white dwarf, a main-sequence star, a giant, or a supergiant.

Astronomers have found subcategories in which to place stars by studying their spectral absorption lines in detail. This refined classification is possible primarily because absorption lines are affected by the density and pressure of the gas in a star's atmosphere, both of which depend on whether the star is a white dwarf, main-sequence star, giant, or supergiant (**Figure 10-9**).

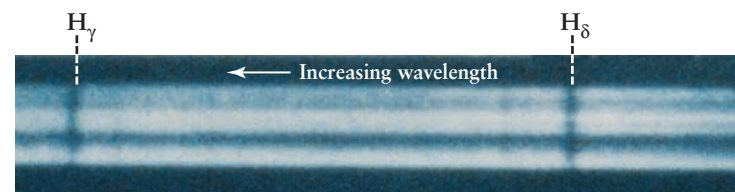
Based upon the differences in stellar spectra, W. W. Morgan, P. C. Keenan, and Edith Kellman of the Yerkes Observatory developed a system of **luminosity classes** in the 1930s. Luminosity classes Ia and Ib encompass all of the supergiants; giants of various luminosity are assigned classes II, III, and IV; and main-sequence stars

Focus Question 10-7

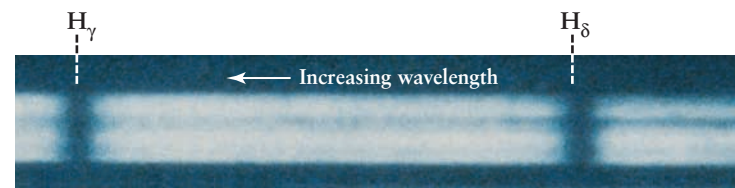
For a given spectral class (that is, temperature), a star of which luminosity class—main sequence, giant, or supergiant—is brightest?

Focus Question 10-8

A main-sequence star of which spectral type—F5, A8, or K0—is largest?



a A supergiant star has a low-density, low-pressure atmosphere: its spectrum has narrow absorption lines



b A main-sequence star has a denser, higher-pressure atmosphere: its spectrum has broad absorption lines

R I V U X G

FIGURE 10-9 Stellar Size and Spectra These spectra are from two stars of the same spectral type (B8) and, hence, the same surface temperature (13,400 K) but different radii and luminosities: **(a)** the B8 supergiant Rigel (luminosity 58,000 $L_{\odot}$) in Orion and **(b)** the B8 main-sequence star Algol (luminosity 100 $L_{\odot}$) in Perseus. (W. W. Morgan, P. C. Keenan, and E. Kellman, *An Atlas of Stellar Spectra*)

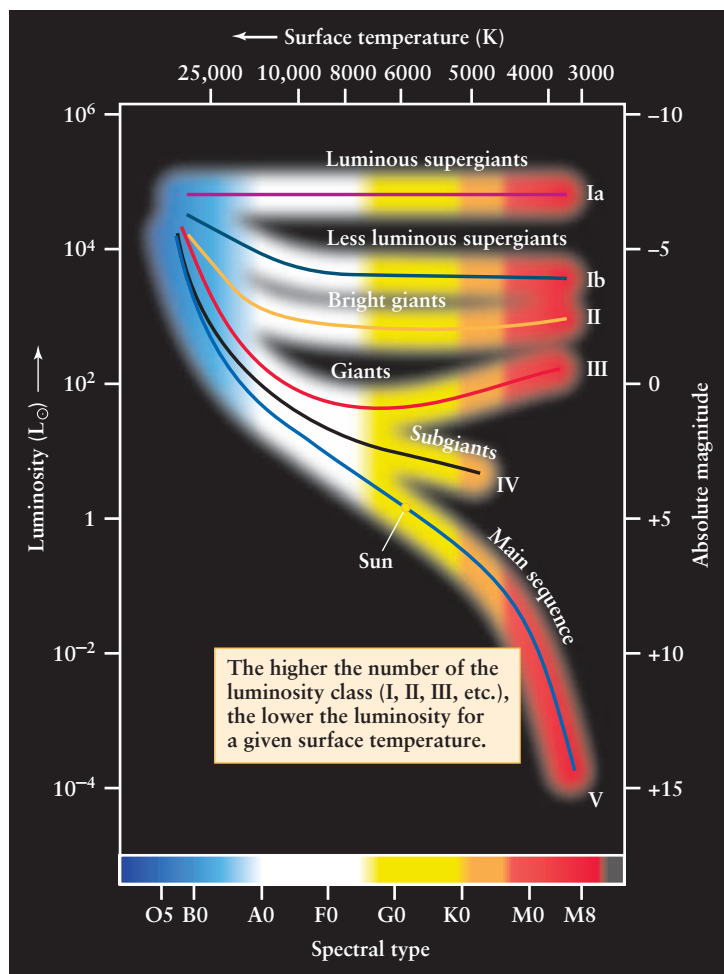


FIGURE 10-10 Luminosity Classes Dividing the H-R diagram into regions, called *luminosity classes*, permits finer distinctions between giants and supergiants. Luminosity classes Ia and Ib encompass the supergiants. Luminosity classes II, III, and IV indicate giants of different brightness. Luminosity class V indicates main-sequence stars. Because they do not create energy by fusion, like the other stars we have discussed so far, white dwarfs do not have their own luminosity class.

are luminosity class V. **Figure 10-10** summarizes these results. We will explore in Chapter 11 how the different luminosity classes correspond to different stages of stellar evolution. Note on Figure 10-10 that white dwarfs are not given a luminosity class because they are stellar remnants that do not create energy by fusion like the Sun and other stars in the five luminosity classes. We will also study white dwarfs further in Chapter 11.

Plotting luminosity classes on the H-R diagram (see Figure 10-10) provides a useful subdivision of star types. In fact, astronomers commonly describe a star by both its spectral type and its luminosity class. The Sun, for example, is a G2 V star. This notation supplies a great deal of information about the star, because its spectral type is correlated with its surface temperature. Thus, an astronomer knows immediately that a G2 V star is a

main-sequence star with a luminosity of $1 L_{\odot}$ and a surface temperature of around 5800 K. Similarly, knowing that Aldebaran is a K5 III star tells an astronomer that it is a red giant with a luminosity of around $370 L_{\odot}$ and a surface temperature of about 4400 K (see Figure 10-7).

10-9 A star's spectral type and luminosity class provide a second distance-measuring technique

A star's spectral type and luminosity class, combined with information from the H-R diagram, give astronomers all the information necessary for determining the distances to stars millions of light-years away, far beyond the maximum distance that can be measured using stellar parallax. The process works like this:

1. Astronomers observe a distant star's apparent magnitude and spectrum.
2. From its spectrum, they determine which spectral class the star belongs to (or, equivalently, its surface temperature).
3. The spectrum also reveals the star's luminosity class, that is, whether it is a supergiant, giant, or main-sequence star. Combining the temperature and the luminosity class determines the star's location on the H-R diagram (see Figure 10-10). From this position, the star's approximate absolute magnitude can be read off the diagram.
4. Using the apparent and absolute magnitudes in the distance-magnitude relationship (see Discovery 10-2), the star's distance can then be calculated.

Consider, for example, the star Regulus in the constellation Leo. Regulus's spectrum reveals it to be a B7 V star (a hot, blue, main-sequence star). Placing it on the H-R diagram (see Figure 10-7), we see that its luminosity is $140 L_{\odot}$ and its absolute magnitude is -0.52 . Given the star's apparent magnitude, we can use the distance-magnitude relationship to determine its distance from Earth. Because both the spectral type and luminosity class are obtained spectroscopically, this method of determining distances is called **spectroscopic parallax**. (This name is misleading, however, because no parallax angle is involved.)

Spectroscopic parallax is limited in accuracy because of the spread of stars in each luminosity class—the stars in each class do not fall on a single, narrow line (hence, the use of the word *approximate* in the preceding numbered process). It is also limited because spectra of distant stars become increasingly hard to determine. As a result, errors of 10% in distance are common using spectroscopic parallax. Accepting such errors, the power of this method is that it can be used for stars at much greater distances than those determined by stellar parallax. Indeed, this method even provides distances to stars in other galaxies tens of millions of light-years away.

STELLAR MASSES

The final clue that astronomers needed to understand how stars work is their mass—how much matter they have. If the masses were related to the stellar luminosities and locations on the H-R diagram (which they are), then astronomers would have enough information to start making scientific models of why stars shine. The mass of each star, it turns out, determines how strongly it can compress and heat its interior and thereby create light and other electromagnetic radiation by thermonuclear fusion, which we will explore in the next chapter. Models of how stars generate that radiation, developed in the twentieth century and still being refined today, provide us with further insights into how nature works. With the knowledge gained about thermonuclear fusion in stars, physicists are researching how to control this fusion to generate electricity.

The problem with finding stellar masses is that it cannot be done directly by examining isolated stars. The mass of a star can be determined only by its gravitational effects on other bodies, using Newton's law of gravity. When Newton derived the formula for Kepler's third law, he discovered that for any two objects in orbit, such as a planet around the Sun, a moon around a planet, or two stars around each other, the period of their orbits is related to the sum of their masses (see **Discovery 10-3: Kepler's Third Law and Stellar Masses**). You have seen in Section 2-5 that for objects orbiting the Sun, we can ignore the mass of the smaller body, thereby getting Kepler's original third law, as stated in Chapter 2. For stars orbiting each other, we use Newton's full equation, which includes the masses of both bodies.

6 Fortunately for astronomers, about one-third of the stars near our solar system are members of star systems in which two stars orbit each other. Put another way, *about one-third of the objects we see as single stars are*

actually pairs of stars in orbit around each other. Most of these pairs of stars are so distant that we cannot resolve the individual stars in them with our naked eyes, and so they appear in the night sky as single objects. However, using telescopes to observe the periods of the orbits and the distances between stars, astronomers can determine the sum of the stellar masses of the pair, and as we will see shortly, they can sometimes also determine individual stellar masses.

Focus Question 10-9

What observations must be made to determine a star's distance using spectroscopic parallax?

10-10 Binary stars provide information about stellar masses

A pair of stars located at nearly the same position in the night sky is called a *double star*. Between 1782 and 1838, William Herschel, his sister Caroline, and his son John catalogued thousands of them. Some double stars are not physically close together and do not orbit each other. These **optical doubles** (sometimes called *apparent binaries*) just happen to lie in the same direction, as seen from Earth. For example, δ Herculis, visible with the naked eye, is an optical double with a dim star. Through a telescope these stars appear close together, but that is an optical illusion.

Other double stars are true **binary stars**—pairs in which two stars orbit each other. In the case of **visual binaries**, both stars can be seen, using a telescope if necessary (**Figure 10-11**). Astronomers can plot the orbit of one star around the other in a visual binary.

To see how we determine the masses of the individual stars, consider a visual binary. Newton rewrote Kepler's third law as a relation between the masses of the stars (in solar masses), their orbital period around each other in Earth years (this period is the same for both stars, so it does not matter which star we assume

DISCOVERY 10-3

Kepler's Third Law and Stellar Masses

The same gravitational force that holds Earth in orbit around the Sun also keeps pairs of stars in orbit around each other. For any pair of orbiting bodies, the orbits are ellipses, and Kepler's third law can be written as

$$M_1 + M_2 = a^3/P^2$$

Here, M_1 and M_2 are the two masses (expressed in solar masses), a is the length of the semimajor axis of the ellipse (in astronomical units), and P is the orbital period (in years). Note that a is also the average separation

between the two bodies. Thus, the sum of the masses can be found once we know the orbital period and average separation.

Example: Suppose two stars make up a binary system, with one star's elliptical orbit having a semimajor axis, a , of 4 AU. We find that this star takes 2.5 years to complete one orbit. Then the sum of the stars' masses is

$$M_1 + M_2 = (4)^3/(2.5)^2 = 10.2 M_\odot$$

The total mass of the system is $10.2 M_\odot$.

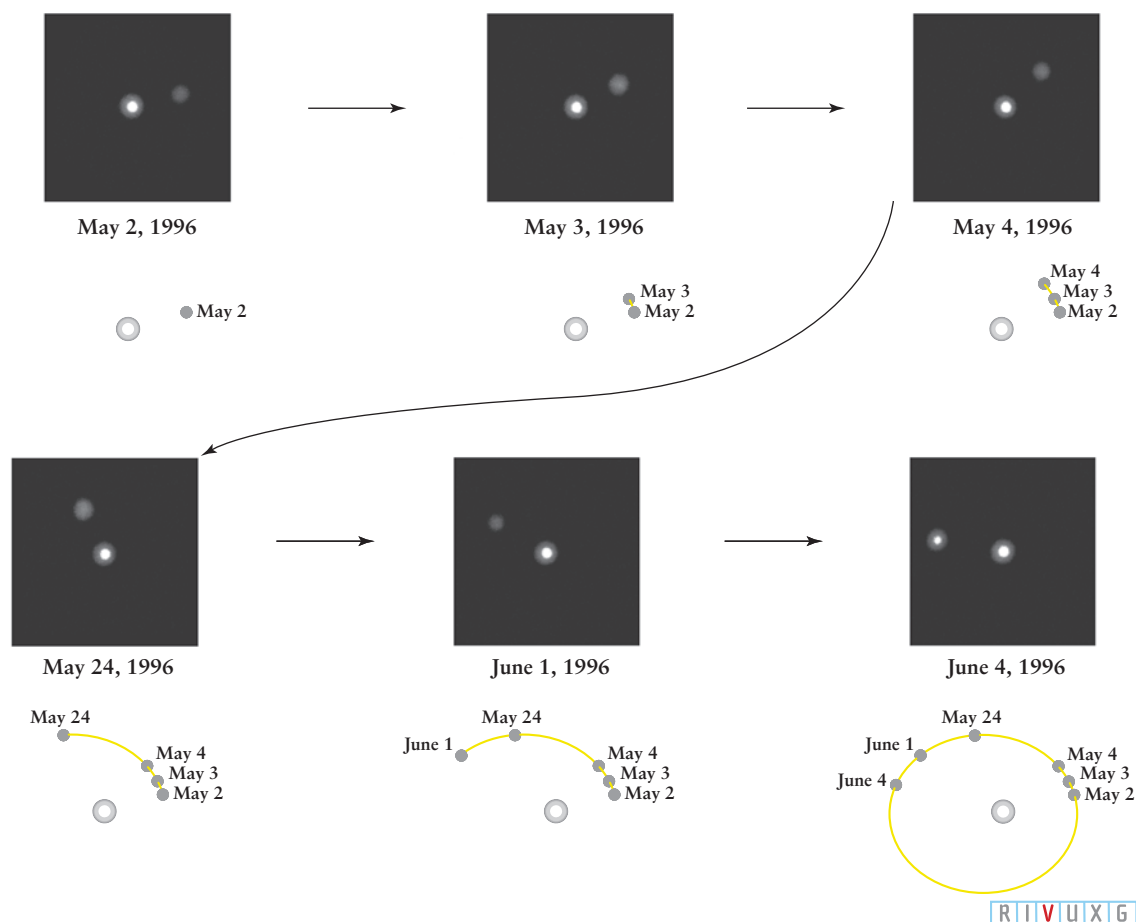


FIGURE 10-11 A Binary Star System About one-third of the objects we see as “stars” in our region of the Milky Way Galaxy are actually double stars. Mizar in Ursa Major is a binary system with stars separated by only about 0.01 arcsec. The images and plots show the relative positions of the two stars over nearly half of their orbital period. The orbital motion of

the two binary stars around each other is evident. Either star can be considered fixed in making such plots. (Technically, this pair of stars is Mizar A and its dimmer companion. These two are bound to another binary pair, Mizar B and its dimmer companion.) (Navy Prototype Optical Interferometer, Flagstaff, AZ, Lowell Observatory)

orbits the other), and the average separation between the stars in AU (which equals the semimajor axis of the elliptical orbit):

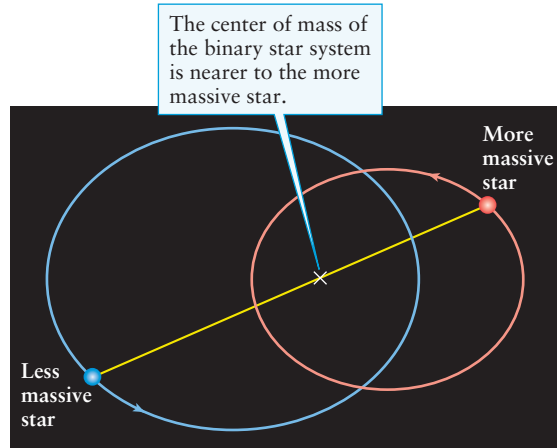
$$\text{the sum of the masses} = \frac{\text{the cube of the semimajor axis}}{\text{the square of the period}}$$

Thus, by observing the separation between a pair of stars and how long one of them takes to complete its orbit, we can calculate the sum of the masses of the two stars, presented mathematically in Discovery 10-3: Kepler’s Third Law and Stellar Masses.

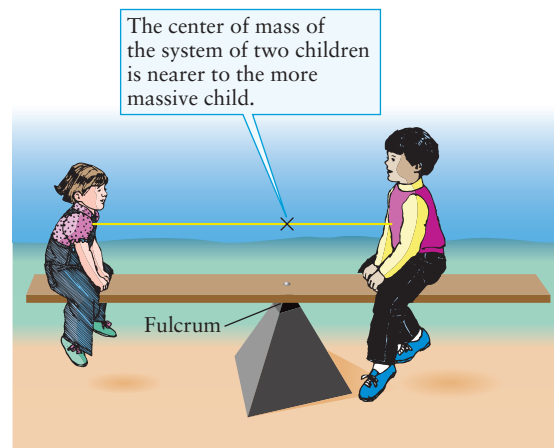
To determine the individual masses of the stars, we need to know the distances of the stars from the center of mass of the pair. As with the Earth–Moon system discussed in Section 6-9, both stars in a binary system move in elliptical orbits around a common point between them. This point, called the **center of mass**, is determined by the masses of the stars and can be understood by analogy to a pair of connected masses whirling on a frictionless surface (see Figure 6-31). One point somewhere

between their centers moves in a straight line. This point is the center of mass of the system of masses. Just as the two masses orbit around their center of mass, the two stars in a binary system orbit around their center of mass under the influence of their mutual gravitational attraction (Figure 10-12). The more massive star is closer to the center of mass, just as a heavier child must be closer to the pivot point or fulcrum of a seesaw for it to balance (Figure 10-12b).

We can locate the center of mass of a visual binary by plotting the separate orbits of the two stars, as in Figure 10-11, using the background stars as reference points. The center of mass lies at the common focus of the two elliptical orbits. To get the relative sizes of the orbits, however, we must be able to determine the plane of orbit of the two stars. This calculation is not always possible, which is why we cannot always determine the individual masses of the stars. However, when the stars in a binary system periodically eclipse each other, as seen from Earth, we know that the plane of



a A binary star system



b A “binary system” of two children

FIGURE 10-12 Center of Mass of a Binary Star System (a) Two stars move in elliptical orbits around a common center of mass. Although the orbits cross each other, the two stars are always on opposite sides of the center of mass and thus never

collide. (b) A seesaw balances if the center of mass of the two children is at the fulcrum. When balanced, the heavier child is always closer to the fulcrum, just as the more massive star is closer to the center of mass of a binary star system.

their orbit must be nearly parallel to our line of sight to them (**Figure 10-13a, b**). Comparing the relative sizes of the two orbits around the center of mass yields the ratio of the two stars’ masses, M_1/M_2 .

When we can determine M_1/M_2 , we can find the individual masses, because we already know the sum $M_1 + M_2$ from Kepler’s third law. The individual masses are determined by combining the ratio of the masses and the sum of the masses (that is, two equations and two unknowns yield the values for both unknowns). The range of stellar masses thus determined extends from

$0.08 M_\odot$ to about $150 M_\odot$. (Stars with masses up to $265 M_\odot$ have been observed by other means.)

Eclipsing binaries can be detected even when the stars cannot be resolved as two distinct images in a telescope. The apparent magnitude of the image of an eclipsing binary dims each time one star blocks out part of the other. An astronomer can measure light intensity from binaries very accurately as a function of time. From these **light curves**, such as those shown in **Figure 10-13**, we can see at a glance whether the eclipse is partial, creating a V-shaped trough in the light curve

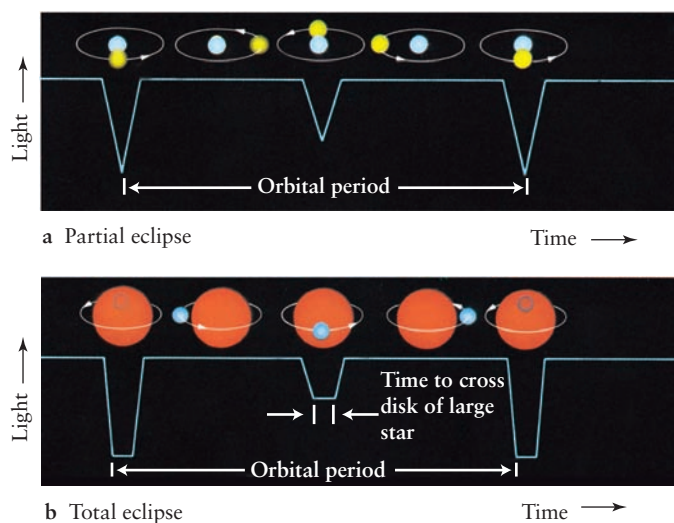
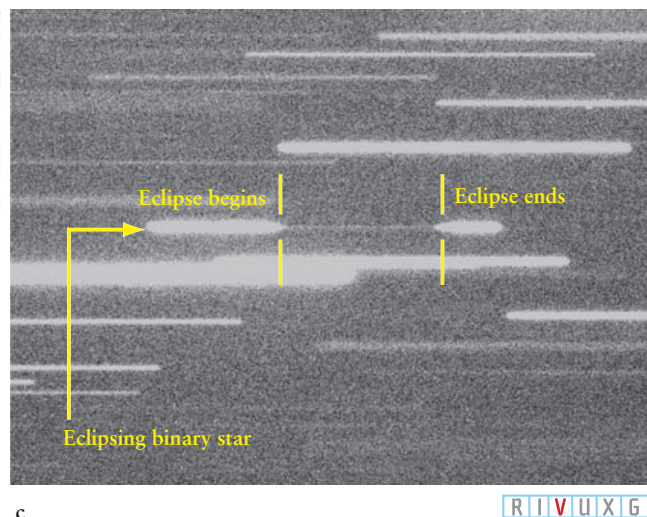


FIGURE 10-13 Representative Light Curves of Eclipsing Binaries The shape of the light curve (blue) reveals that the pairs of stars have orbits in planes nearly edge-on to our line of sight. It also provides details about the two stars that make up an eclipsing binary. Illustrated here are (a) a partial eclipse and (b) a total eclipse. (c) The binary star NN Serpens, indicated by



the arrow, undergoes a total eclipse. The telescope was moved during the exposure so that the sky drifted slowly from left to right. During the 10.5-minute eclipse, the dimmer but larger star in the binary system (an M6 V star) passed in front of the more luminous but smaller star (a white dwarf). The binary became so dim that it almost disappeared. (c: European Southern Observatory)

(Figure 10-13a), or total, creating a flat-bottomed trough (Figure 10-13b). These data provide details of how close the plane of their orbit is to being perpendicular to our line of sight.

Light curves can also reveal information about stellar atmospheres. Suppose that one star of a binary is a tiny white dwarf (the size of Earth) and the other is a giant (much larger than the Sun). By observing exactly how the light from the white dwarf is gradually cut off as it begins to move behind the edge of the giant during an eclipse, astronomers can infer the pressure and density in the upper atmosphere of the giant. Such information is invaluable in testing models of stellar structure.

Many binary stars are separated by several astronomical units or more. Other than orbiting each other, these stars behave as though they are isolated; that is, the models we develop in the following chapters for the evolution of isolated stars apply to these binary stars as well. However, there are also **close binary** systems, with only a few stellar diameters separating the stars. Such stars are so close together that the gravity of each one dramatically affects the appearance and evolution

of the other. If one member of a close binary is a giant, some of the gas of its outer layers is pulled onto its more compact companion—mass is transferred from one star to the other. We will explore more about such systems in Chapter 11. In addition to binary systems of stars, a few systems composed of three stars bound to each other exist in our neighborhood of space. Polaris, the North Star, for example, is part of a triple system.

10-11 Main-sequence stars have a relationship between mass and luminosity

As binary star data accumulated, an important trend began to emerge. On the main sequence, the more luminous the star, the more massive it is. This mass–luminosity relationship can be conveniently displayed graphically (Figure 10-14a); note that the axes of this figure are logarithmic. The Sun’s mass lies in the middle of both logarithmic ranges. The **mass–luminosity relationship** demonstrates that *the main sequence on the H-R diagram is a progression in mass as well as in luminosity and surface temperature*. The hot, bright, bluish stars in the upper

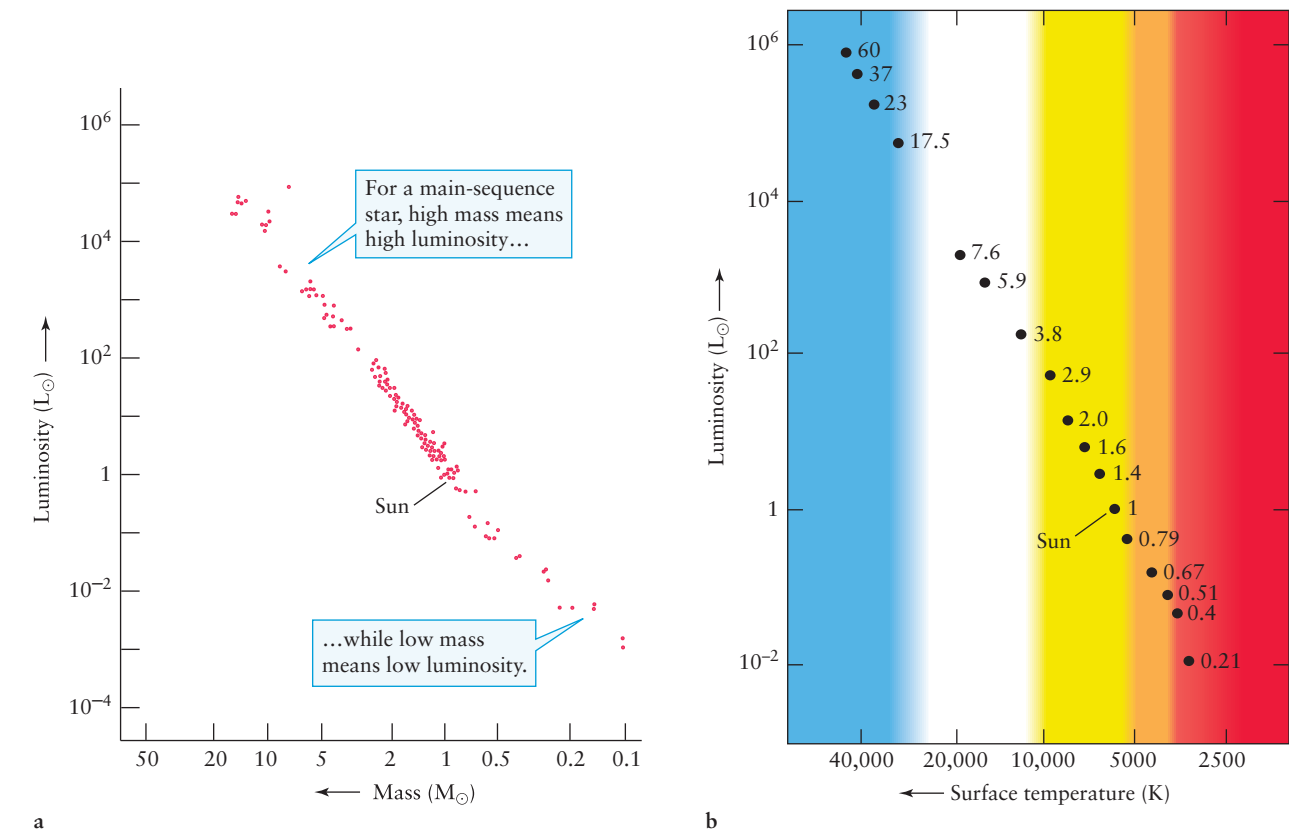


FIGURE 10-14 The Mass–Luminosity Relationship (a) For main-sequence stars, mass and luminosity are directly correlated—the more massive a star, the more luminous it is. A main-sequence star of mass $10 M_{\odot}$ has roughly 3000 times the Sun’s luminosity ($3000 L_{\odot}$); one with $0.1 M_{\odot}$ has a luminosity of only about $0.001 L_{\odot}$. To fit the whole sequence on one page, the

luminosities and masses are plotted using logarithmic scales. (b) Equivalently, on this H-R diagram, each dot represents a main-sequence star. The number next to each dot is the mass of that star in solar masses ($M_{\odot}$). As you move up the main sequence from the lower right to the upper left, the mass, luminosity, and surface temperature of main-sequence stars all increase.

left corner of the H-R diagram (see Figure 10-14b) are the most massive main-sequence stars. As we move down the sequence, stellar masses decrease until we reach the dim, cool, reddish stars in the lower right corner of the H-R diagram. The correlation of stellar luminosities with mass suggests (correctly, as it turns out) that a star's energy production is closely linked to its mass. Indeed, mass is *the* crucial factor in explaining why stars are as bright as they are and how they evolve.

10-12 The orbital motion of binary stars affects the wavelengths of their spectral lines

Many binary stars are scattered throughout our Milky Way Galaxy, but only those that are nearby or are widely separated can be distinguished as visual binaries. A remote binary often presents the appearance of a single star because even our best telescopes cannot resolve the images of its individual stars. These binary systems can still be detected, however, through spectroscopy.

Spectral analysis yields incongruous spectral lines for some stars. For example, the spectrum of what appears at first to be a single star may include strong absorption lines for both hydrogen (indicating a hot, type A star) and titanium oxide (indicating a cool, type M star). Because a single star cannot display both types of absorption lines prominently, such an observation must be revealing a binary system.

Spectroscopy can also detect the movements of stars orbiting each other because of the Doppler shift in spectral lines. As we saw in Section 3-2, an approaching source of light has shorter wavelengths than if the same source were stationary, and a receding source has longer wavelengths. The amount of the shift is proportional to the speed of the light source: the greater the speed, the greater the shift.

If the two stars in a binary are orbiting at more than a few kilometers per second, they produce spectral lines that shift back and forth regularly. Such stars are called **spectroscopic binaries**. The motions of the stars revolving around their center of mass produce the periodic shifting of the spectral lines. In many spectroscopic binaries, one of the stars is so dim that its spectral lines cannot be detected. Instead, a single set of spectral lines from the star that we can see shifts regularly back and forth. Such a *single-line spectroscopic binary* yields less information about its two stars than does a *double-line spectroscopic binary*, in which the lines from both stars are visible. In particular, a double-line binary, in which one star eclipses the other, yields the individual masses of its two stars, whereas a single-line binary cannot provide that information.

Figure 10-15 shows four spectra of a spectroscopic binary system. At Stage 1, star A is moving toward

Earth; hence, its spectrum is blueshifted. At the same time, star B is moving away, so its spectrum is redshifted. In the spectrum below the drawing, you can see two sets of spectral lines, each slightly offset from each other. A few days later, at Stage 2, the stars have progressed along their orbits so that star A is moving toward the right and star B toward the left, as seen from Earth. Because neither star is moving toward or away from us, we see no Doppler shift and so their spectral lines return to their normal positions. Stage 3 shows the opposite spectra from Stage 1 (star A is now redshifted and star B is blueshifted), while Stage 4 shows the same spectrum as Stage 2. **Figure 10-16** shows spectra taken of the binary star system κ (kappa) Arietis, demonstrating these shifts.

The shifts in spectral lines can yield significant information about the orbital velocities of the stars in a spectroscopic binary. This information is best displayed as a **radial-velocity curve**, in which radial velocity is graphed over time. In Figure 10-15, the wavy pattern repeats with a period of about 15 days, which is the orbital period of the binary. This pattern is displaced upward from the zero-velocity line by about 12 km/s, which is the overall motion of the binary system away from Earth. Superimposed on this overall recessional motion are the periodic approaches and recessions of the two stars as they orbit around each other.

Combining radial-velocity curves and light curves, we can obtain other useful information from eclipsing binaries. If an eclipsing binary is also a double-line spectroscopic binary, astronomers can calculate the masses, diameters, varied brightness, speeds, and stellar separation of each star. These stars are rare, however, because the orbital planes of most spectroscopic binaries are tilted so that eclipses do not occur, as seen from Earth.

While most of the stars we see in the night sky are single objects or in binary systems, some are in groups of three or more. For example, the Pleiades in the constellation Taurus is a cluster of hundreds of stars, of which seven (the “seven sisters”) are often visible to the naked eye. Indeed, astronomers observe that many stars throughout our Milky Way Galaxy and elsewhere are found in even larger groups, which we will study in the next chapter.

This chapter has provided us with observational data that serve, in large measure, as the basis for the models of stellar activity and evolution presented in the next two chapters. The models make predictions that lead to new observations and refined models. As we proceed through these models, we will take up some further observations driven by these refinements.

Focus Question 10-10

For what types of stars does the mass–luminosity relationship not apply?

Focus Question 10-11

How do the spectral lines of a spectroscopic binary change as observed from Earth?

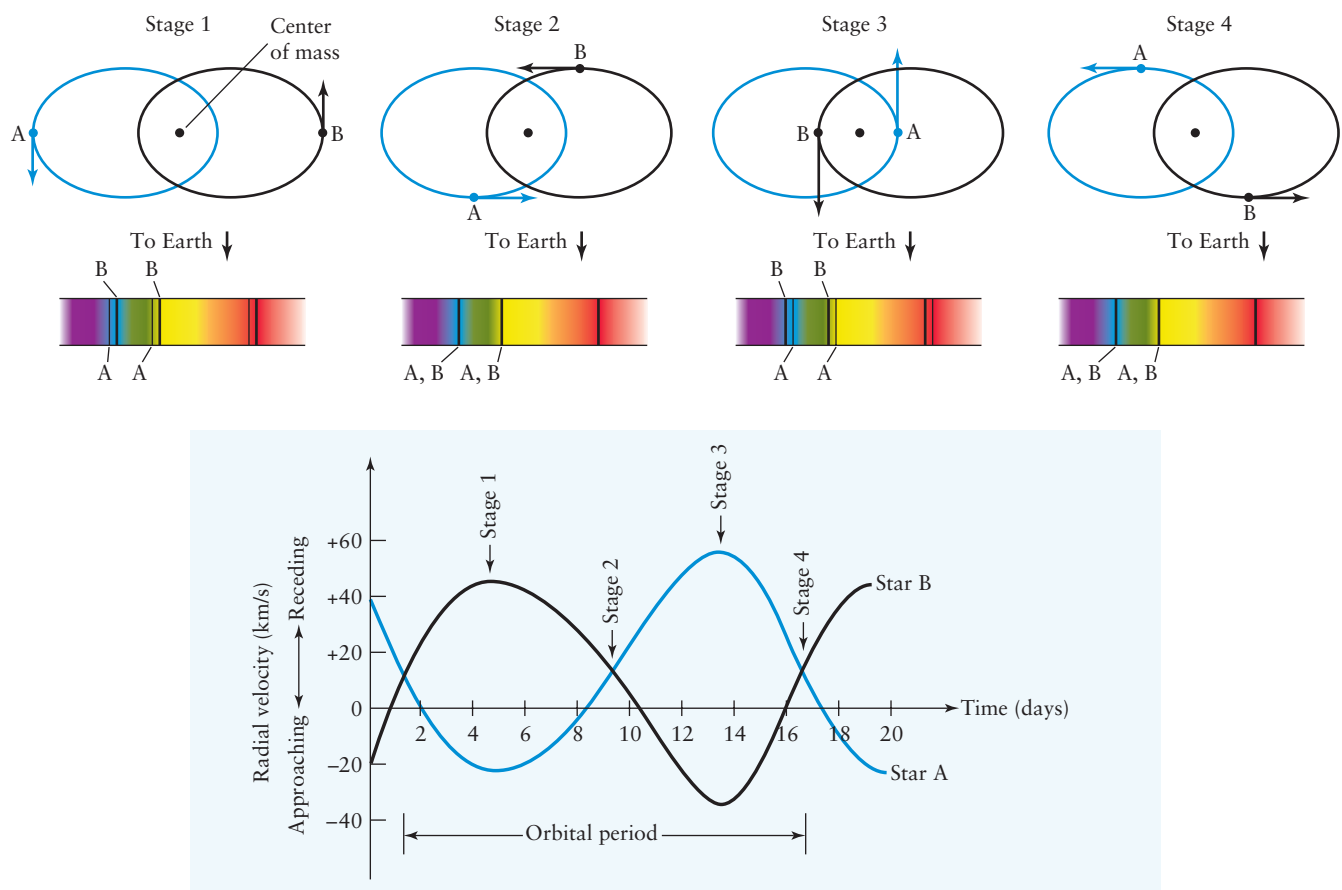


FIGURE 10-15 Spectral-Line Motion in Binary Star Systems The diagrams at the top indicate the positions and motions of the stars, labeled A and B, relative to Earth. Below each diagram is the spectrum we would observe for these two stars at each stage. The changes in colors (wavelengths) of the spectral lines are due to changes in the stars' Doppler shifts, as

seen from Earth. The graph displays the radial-velocity curves of the binary HD 171978. (The HD means that this is a star from the *Henry Draper Catalogue* of stars.) The entire binary is moving away from us at 12 km/s, which is why the pattern of radial-velocity curves is displaced upward from the zero-velocity line.

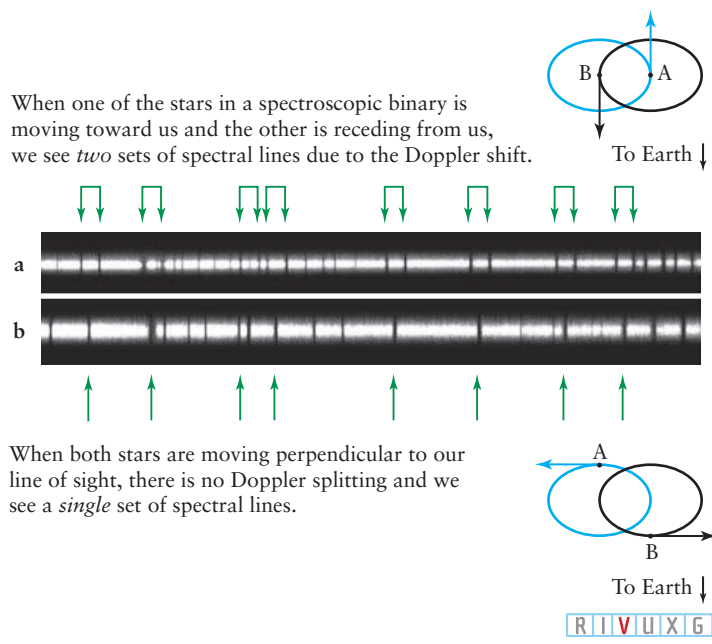


FIGURE 10-16 A Double-Line Spectroscopic Binary The spectrum of the double-line spectroscopic binary κ (kappa) Arietis has spectral lines that shift back and forth as the two stars revolve around each other. (a) The stars are moving parallel to the line of sight, with one star approaching Earth, the other star receding, as in Stage 1 or 3 of Figure 10-15. These motions produce two sets of shifted spectral lines. (b) Both stars are moving perpendicular to our line of sight, as in Stage 2 or 4 of Figure 10-15. As a result, the spectral lines of the two stars have merged.

SUMMARY OF KEY IDEAS

• Stars differ in size, luminosity, temperature, color, mass, and chemical composition—facts that help astronomers understand stellar structure and evolution.

Magnitude Scales

- Determining stellar distances from Earth is the first step to understanding the nature of the stars. Distances to the nearer stars can be determined by stellar parallax, which is the apparent shift of a star's location against the background stars while Earth moves along its orbit around the Sun. The apparent magnitude of a star, denoted m , is a measure of how bright the star appears to Earth-based observers. The absolute magnitude of a star, denoted M , is a measure of the star's true brightness and is directly related to the star's energy output, or luminosity.
- The luminosity of a star is the amount of energy it emits each second.
- The absolute magnitude of a star is the apparent magnitude it would have if viewed from a distance of 10 pc (32.6 ly). Absolute magnitudes can be calculated from the star's apparent magnitude and distance from Earth.

The Temperatures of Stars

- Stellar temperatures can be determined from stars' colors or stellar spectra.
- Stars are classified into spectral types (O, B, A, F, G, K, and M) based on their spectra or, equivalently, their surface temperatures.

Types of Stars

- The Hertzsprung-Russell (H-R) diagram is a graph on which luminosities of stars are plotted against their spectral types (or, equivalently, their absolute magnitudes are plotted against surface temperatures).
- The H-R diagram reveals the existence of four major groupings of stars: main-sequence stars, giants, supergiants, and white dwarfs.
- The mass-luminosity relationship expresses a direct correlation between a main-sequence star's mass and the total energy it emits.
- Distances to stars can be determined using their spectral types and luminosity classes, a method called spectroscopic parallax.

Stellar Masses

- Binary stars are fairly common. Those that can be resolved into two distinct star images (even if it takes a telescope to do this) are called visual binaries.
- The masses of the two stars in a binary system can be computed from measurements of the orbital period and orbital dimensions of the system.

- Some binaries can be detected and analyzed, even though the system may be so distant (or the two stars so close together) that the two star images cannot be resolved with a telescope.
- A spectroscopic binary is a system detected from the periodic shift of its spectral lines. This shift is caused by the Doppler effect as the orbits of the stars carry them alternately toward and away from Earth.
- An eclipsing binary is a system whose orbits are viewed nearly edge-on from Earth, so that one star periodically eclipses the other. Detailed information about the stars in an eclipsing binary can be obtained by studying the binary's light curve.

WHAT DID YOU THINK?

- How near to us is the closest star other than the Sun?* The closest star, Proxima Centauri, is about 40 trillion km (25 trillion mi) away. Light from there takes about 4 years to reach Earth.
- How luminous is the Sun compared with other stars?* The most luminous stars are about a million times brighter, and the least luminous stars are about a hundred thousand times dimmer than the Sun.
- What colors are stars, and why do they have these colors?* Stars are found in a wide range of colors, from red through violet as well as white. They have these colors because they have different surface temperatures.
- Are brighter stars hotter than dimmer stars?* Not necessarily. Many brighter stars (such as red giants) are cooler but larger than hotter, dimmer stars (such as white dwarfs).
- Compared to the Sun, what sizes are other stars?* Stars range from more than 1000 times the Sun's diameter to less than 1/100 of the Sun's diameter.
- Are most stars isolated from other stars, as the Sun is?* Yes. In the vicinity of the Sun, one-third of stars are found in gravitationally bound pairs or larger groups.

Review Questions

1. Stellar parallax measurements are used in astronomy to determine which of the following properties of stars?

- | | |
|-------------------|-----------------|
| a. sizes | d. colors |
| b. rotation rates | e. temperatures |
| c. distances | |

2. The brightness a star would have if it were at 10 pc from Earth is called its

- | | |
|-----------------------|-------------------|
| a. absolute magnitude | d. spectral type |
| b. apparent magnitude | e. center of mass |
| c. luminosity | |

3. Measurements of a binary star system are required to determine what property of stars?
 - a. luminosity
 - b. apparent magnitude
 - c. distance
 - d. mass
 - e. temperature
4. A star with which of the following apparent magnitudes appears brightest from Earth?
 - a. 6.8
 - b. 3.2
 - c. 0.41
 - d. -0.44
 - e. -1.5
5. A star of what spectral class has the strongest (darkest) H_{α} line? (*Hint: See Figure 10-5.*)
 - a. B2
 - b. A0
 - c. A5
 - d. G5
 - e. M0
6. Describe how the parallax method of finding a star's distance is similar to the binocular (two-eye) vision of animals.
7. What is stellar parallax?
8. How do astronomers use stellar parallax to measure the distances to stars?
9. Why do stellar parallax measurements work only with relatively nearby stars?
10. What is the difference between apparent magnitude and absolute magnitude?
11. Briefly describe how you would determine the absolute magnitude of a nearby star.
12. What does a star's luminosity measure?
13. How is the magnitude scale "backward" from what common sense dictates?
14. Does the star Betelgeuse, whose apparent magnitude is $m = +0.5$, look brighter or dimmer to us than the star Pollux, whose apparent magnitude is $m = +1.1$?
15. How and why is the spectrum of a star related to its surface temperature?
16. What is the primary chemical component of most stars?
17. A star of which spectral type has the strongest Na I absorption lines? At approximately what wavelength is this line normally found? (*Hint: See Figure 10-5.*)
18. Why does a G2 star have many more absorption lines than a B0 star?
19. Draw an H-R diagram and sketch the regions occupied by main-sequence stars, giants, supergiants, and white dwarfs. Briefly discuss the different ways you could have labeled the axes of your graph.
20. To test your understanding of the H-R diagram, do Interactive Exercise 11.1 on the assigned Web site. You can print out your answers, if required.
21. How can observations of a visual binary lead to information about the masses of its stars?
22. What is a radial-velocity curve? What kinds of stellar systems exhibit such curves?
23. What is the difference between a single-line and a double-line spectroscopic binary?
24. What is meant by the light curve of an eclipsing binary? What sorts of information can be determined from such a light curve?
25. What is the mass–luminosity relationship? To what kind of stars does it apply?
26. Referring to Figure 10-7:
 - a. Which are the hottest and coolest named stars on the diagram?
 - b. Which are the brightest and dimmest named stars on the diagram?
 - c. Which are the hottest and coolest named main-sequence stars on the diagram?
 - d. Which named stars are white dwarfs? Giants? Supergiants?

Advanced Questions

27. What is the inverse-square law? Use it to explain why a headlight on a car can appear brighter than a star, even though the headlight emits far less light energy per second.
28. Explain how sailors on a ship traveling parallel to a coastline at a known speed can use parallax angle measurements to determine the distance to the shore.
29. Sketch the radial-velocity curve of a binary whose stars are moving in nearly circular orbits that are (a) perpendicular and (b) parallel to our line of sight.
30. Sketch the light curve of an eclipsing binary whose stars are moving along highly elongated orbits with the major axis of the orbits (a) pointed toward Earth and (b) perpendicular to our line of sight.

Discussion Questions

31. Discuss the advantages and disadvantages of measuring stellar parallax from a space telescope at the distance of Jupiter from the Sun compared to the same measurements made from Earth's surface.
32. How does a star's rotation affect the appearance of its spectral lines? *Hint: Assume we are looking at the star from above its equator. Then, at every instant, half of the spinning star is approaching Earth, while the other half is receding. Consider the resulting Doppler shift in the spectral lines.*

Got It?

33. Of all the spectral classes of stars (O, B, A, F, G, K, M), which is least common?

34. Comparing them as if they were all the same distance from us, which of the following statements most accurately describes the Sun's luminosity relative to other stars?

- a. It is among the most luminous of stars.
- b. It is of average luminosity.
- c. It is less luminous than the average luminosity of stars.

- d. It is among the least luminous of all stars.
- e. All stars are equally luminous.

35. What colors are the hottest and coolest stars?

36. If a star is a red giant, how does its surface temperature compare to that of the Sun?

This page intentionally left blank



R I V I U X I G The young superstar cluster R136 is located in the Tarantula Nebula in the nearby Large Magellanic Cloud galaxy. These stars, surrounded by glowing interstellar clouds, are only a few million years old. (NASA, ESA, and F. Paresce [INAF-IASF, Bologna, Italy], R. O'Connell [University of Virginia, Charlottesville], and the Wide Field Camera 3 Science Oversight Committee)

The Lives of Stars from Birth Through Middle Age

WHAT DO YOU THINK?

- 1 How do stars form?
- 2 Are stars still forming today? If so, where?
- 3 Do more massive stars shine longer than less massive ones? Explain your reasoning.
- 4 When stars like the Sun stop fusing hydrogen into helium in their cores, do the stars get smaller or larger?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

Gravity provides the energy that enables stars to shine. As we saw in studying the Sun (Chapter 9), gravity compresses stars, thereby heating them and causing fusion to occur deep within them. As a result, stars emit huge amounts of radiation. Over time, their chemical compositions, masses, and brightnesses all vary—stars age. Major stages in the life of each star can last from days to hundreds of billions of years. Cumulative lifetimes of stars range from millions to hundreds of billions of years. We will explore their evolution in this and the next chapter.

In this chapter you will discover

- how stars form
- what a stellar “nursery” looks like
- how astronomers use the physical properties of stars to learn about stellar evolution
- the remarkable transformations of older stars into giants
- how the Hertzsprung-Russell (H-R) diagram is your guide to the stellar life cycle
- how pairs of orbiting stars can change each other

Insight Into Science

Beware Poetic License All too often, the names and descriptions we use for nonliving objects seem to give them human qualities. Anthropomorphism—assigning human attributes to nonhuman creatures or even to nonliving objects, like stars—is descriptive but not scientific. In fact, scientists avoid this practice where possible because it creates unjustified expectations. Nevertheless, anthropomorphism does creep into astronomy. For lack of better terms, astronomers describe stellar change with words related to living things, such as *life*, *life cycle*, *evolution*, *birth*, *aging*, *maturing*, *growing old*, and *dying*. This is poetic license of a sort: It piques our interest but muddles our science. As you study stellar evolution, always be mindful of that.

PROTOSTARS AND PRE-MAIN-SEQUENCE STARS

We saw in Chapter 4 that the solar system is believed to have formed from a collapsing, rotating fragment of a cloud of interstellar gas and dust some 4.6 billion years ago. This scientific theory requires that interstellar matter existed before then so that the solar system had something from which to coalesce. Taking this reasoning one step further, if such interstellar matter exists today, perhaps star formation continues.

11-1 Gas and dust exist between the stars

Matter does indeed exist between the stars. We can see a relatively small amount of it through visible-light telescopes (Figure 11-1a); however, the bulk of it is too cold to be seen optically and requires the use of infrared (Figure 11-1b) and radio telescopes. We call all the matter between stars the **interstellar medium**, and it contains at least 10% of the observed mass in our Galaxy. Observations of the spectra of the interstellar medium reveal that it is composed of gas containing isolated atoms and molecules and tiny pieces of dust. Table 11-1 summarizes the composition of the interstellar medium.

More than 200 different types of molecules have been discovered in interstellar space, some with as many as 60 molecules bonded together. Among these are molecular hydrogen (H_2), carbon monoxide (CO), carbon

dioxide (CO_2), water (H_2O), ammonia (NH_3), formaldehyde (H_2CO), the sugar glycoaldehyde ($\text{C}_2\text{H}_4\text{O}_2$), and the 60-carbon-atom buckminsterfullerene (commonly called buckyballs). We can identify these molecules by their unique spectral emissions, just as we can identify elements from atomic spectra.

The dust found in space also has a variety of structures. Some carbon-based (that is, organic) particles out there are typically 0.005 micron (μm) across (10,000 times smaller than the diameter of a typical human hair). Much larger pieces of space dust have cores of carbon or silicon compounds surrounded with mantles of ice and other materials. These grow to more than 0.30 μm in diameter (still nearly 200 times smaller than the diameter of a human hair).

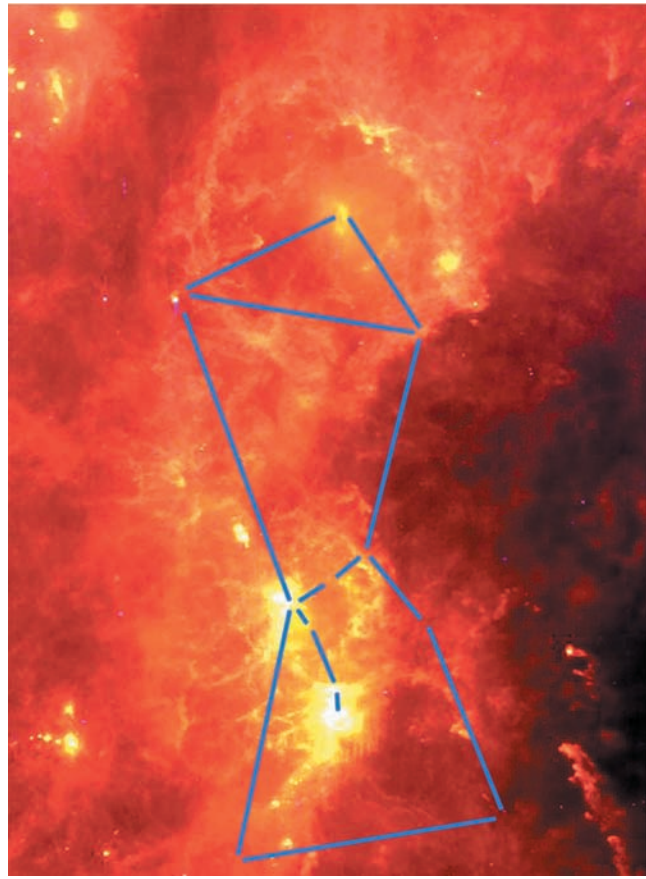
Working with the knowledge that new stars form from the gas and dust of the interstellar medium, astronomers map this matter to identify places to look for young stars. When it can be seen in the visible part of the spectrum, the gas and dust in the interstellar medium glow as a result of scattered light from stars in the vicinity, as seen in Figure 11-1a. As another example, the interstellar medium is dramatically highlighted by stars in the Pleiades star cluster (Figure 11-2a) located in the constellation Taurus. The bluish haze, starlight scattered by the interstellar gas and dust, is called a **reflection nebula**. A **nebula** (plural, *nebulae*) is a dense region of interstellar gas and dust. Nebulae are often embedded



a

R I V U X G

FIGURE 11-1 Stars and the Interstellar Medium (a) The stars in the constellation Orion can easily be seen with the naked eye (see also Figure 1-4). A careful examination of the middle “star” in the “sword” of Orion (which hangs down below his belt) shows that it is not circular. This is actually the center of the Orion Nebula (also called the Great Nebula of Orion). This central region of the Orion Nebula is composed of several stars (the Trapezium) and surrounding gas and dust. This middle “star” in the sword appears fuzzy due to visible light



b

R I V U X G

emitted by a cloud of gas and dust surrounding the Trapezium stars there. Little other interstellar gas and dust can be seen in this visible-light image, although the leftmost star in Orion’s belt, Alnitak, is surrounded by gas and dust (see Figure 11-3). (b) The same region of the sky in a false-color infrared image taken by the Infrared Astronomical Satellite. Gases are seen here to exist in more areas than can be detected in visible light. (a: Akira Fujii/David Malin Images; b: Infrared Astronomical Satellite/NASA/NIVR/SERC)

in much larger bodies of gas and dust, called **molecular clouds**. The largest molecular clouds, called **giant molecular clouds**, contain upward of a million solar masses of matter and extend up to about 600 light-years across.

TABLE 11-1 COMPOSITION OF THE INTERSTELLAR MEDIUM

	Particle number (%)	Mass (%)
Hydrogen (atoms and molecules)	90	74
Helium	9	25
Metals*	1	1

*In astronomy, all elements except hydrogen and helium are called metals.

Astronomers identify several other types of nebulae. Most important among these are dark nebulae and emission nebulae. **Dark nebulae** (Figure 11-2b) are regions of interstellar gas and dust that are sufficiently dense to prevent most of the visible light from behind them from getting to us. They look like regions of empty space, whereas they are actually among the densest of interstellar nebulae. **Emission nebulae** are regions of interstellar gas and dust that glow from energy they receive from nearby stars, from exploding stars, and from collisions between nebulae (Figure 11-3).

While the effect of a dark nebula is an extreme case, finding more distant gas and dust in space is always difficult because the nearby interstellar medium dims, or even blocks, light from behind it. The same effect occurs on Earth where thin cloud layers in our night sky dim

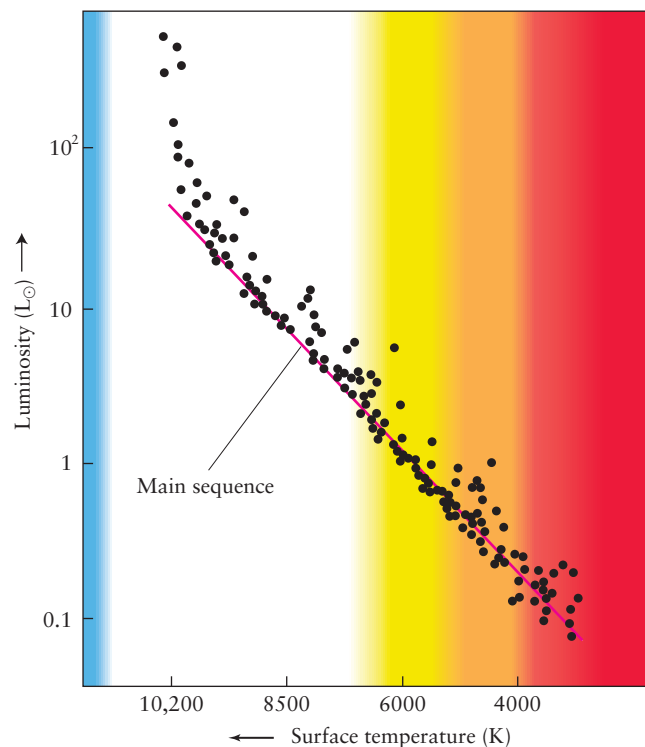
Focus Question 11-1

The interstellar medium is composed primarily of what kinds of things?

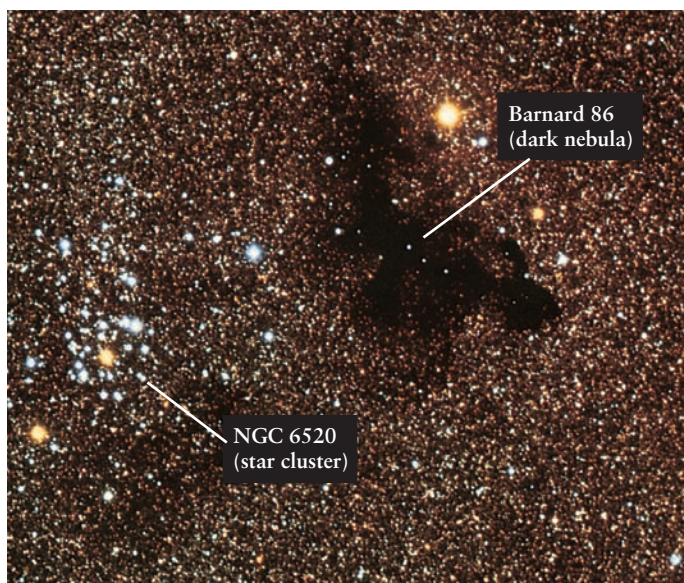


a

R I V U X G



b



c

R I V U X G

FIGURE 11-2 A Reflection Nebula and Dark Nebula

(a) This open cluster, called the Pleiades, can easily be seen with the naked eye in the constellation Taurus (the Bull). Pleiades lies about 440 light-years (134 pc) from Earth. The stars are not shedding mass, unlike the stars in Figures 11-13a and 11-21. The blue glow surrounding the stars of the Pleiades is a reflection nebula created as some of the stars' radiation scatters off preexisting dust grains in their vicinity. (b) Each dot plotted on this H-R diagram represents a star in the Pleiades whose luminosity and surface temperature have been determined. Note that most of the cool, low-mass stars have arrived at the main sequence, indicating that hydrogen fusion has begun in their cores. The cluster has a diameter of about 5 light-years, is about 100 million years old, and contains about 500 stars. (c) The dark nebula Barnard 86 is located in Sagittarius. It is visible in this photograph simply because it blocks out light from the stars beyond it. The bluish stars to the left of the dark nebula are members of a star cluster called NGC 6520. (a and c: Australian Astronomical Observatory/David Malin Images)

the starlight and even prevent us from seeing the stars behind them. This darkening of light by intervening gas and dust in space is called **interstellar extinction**. Dark nebulae are extreme examples of interstellar extinction.

Even when we can see a star or other object at visible wavelengths through the interstellar medium, it appears redder than it actually is, a phenomenon called **interstellar reddening**. This effect occurs because short-wavelength starlight is scattered by dust grains more than is long-wavelength light. (Violet is scattered most, followed by blue, green, yellow, and orange, with red scattered least.)

Therefore, when we observe a star through gas and dust, we are seeing less short-wavelength light from it than we would if the interstellar medium were not there (Figure 11-4a). The more dust between the objects and us, the redder the objects appear (Figure 11-4b). Interstellar reddening is different from reddening due to the Doppler shift (see Section 3-2). The Doppler shift causes all wavelengths of electromagnetic radiation to lengthen equally, whereas interstellar reddening, due to the stronger scattering of shorter wavelengths, does not change the wavelengths of the starlight we receive—only their intensities.

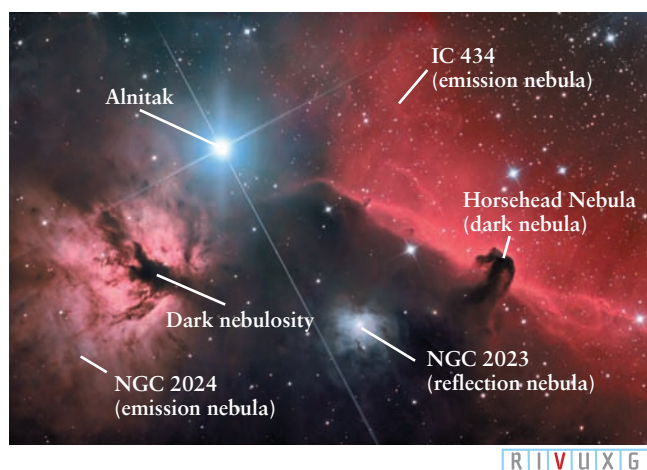


FIGURE 11-3 A Gas- and Dust-Rich Region of Orion A variety of nebulae appear in the sky around Alnitak, also called ζ (zeta) Orionis, the easternmost star in the belt of Orion. To the left of Alnitak is a bright, red emission nebula, called NGC 2024. The glowing gases in emission nebulae are excited by ultraviolet radiation from young, massive stars. Dust grains obscure part of NGC 2024, giving the appearance of black streaks, while the distinctively shaped dust cloud, called the Horsehead Nebula, blocks the light from the background nebula IC 434. The Horsehead Nebula is part of a larger complex of dark interstellar matter, seen in the bottom center of this image. To the left of the Horsehead Nebula is the reflection nebula NGC 2023, whose dust grains scatter blue light from stars between us and it more effectively than any other color. All of this nebulosity lies about 1600 light-years from Earth, while the star Alnitak is only 815 light-years away from us. NGC refers to the New General Catalog of stars and IC stands for Index Catalogs, two supplements to the NGC. (Stocktrek Images/Getty Images)

We have been able to discover distant stars and clouds of interstellar gas and dust whose visible light is

obscured by the nearby interstellar medium because the distant objects also emit radio or infrared photons that are scattered relatively little on their way to us. Likewise, we use radio and infrared telescopes to find nearby interstellar gas and dust that, as blackbodies, are too cold to emit much visible light.

Stars form from gas and dust that become **Jeans unstable** (see Section 4-2). Most of this matter is hydrogen in the form of molecules (rather than atoms). However, molecular hydrogen is relatively hard to detect in space. Therefore, radio astronomers often search for cold interstellar gas in the form of carbon monoxide (CO), which emits lots of photons at short radio (microwave) wavelengths of 2.6 and 1.3 mm. Calculations based on the known abundances of interstellar elements reveal that there are about 10,000 hydrogen molecules (H_2) for every CO molecule in the interstellar medium. Consequently, wherever astronomers detect strong emission of CO, they deduce that an enormous amount of molecular hydrogen gas must also be present.

In mapping the locations of CO emission, astronomers came to realize that interstellar gas and dust are often concentrated in giant molecular clouds. In some cases, these clouds appear as dark nebulae silhouetted against a glowing background light, such as Orion's famous Horsehead Nebula (Figure 11-3). In other cases, the clouds appear as dark nebulae that obscure the background stars (Figure 11-2b). Some 6000 giant molecular clouds are estimated to exist in our Milky Way Galaxy and have masses that range from 10^5 to $2 \times 10^6 M_\odot$ with diameters that range from 15 to 600 light-years. The density inside each of these clouds ranges from 10^2 to 10^5 hydrogen molecules per cubic centimeter—thousands of times greater than the average density of the gas and dust dispersed throughout interstellar space, but some 10^{15} times less dense than the air we breathe.

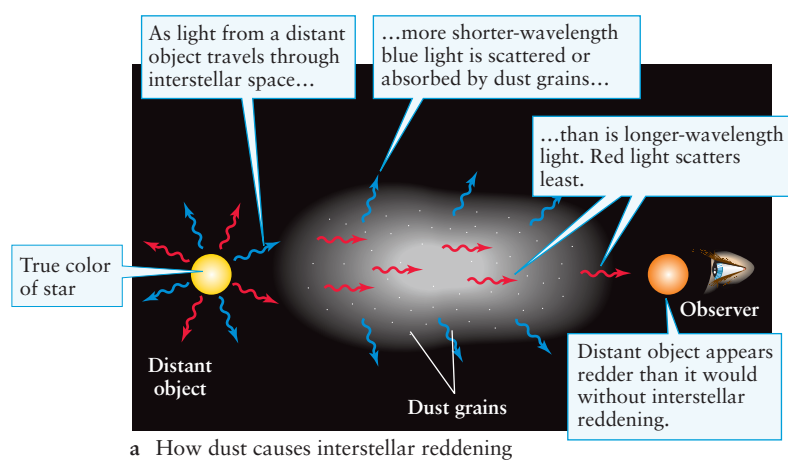
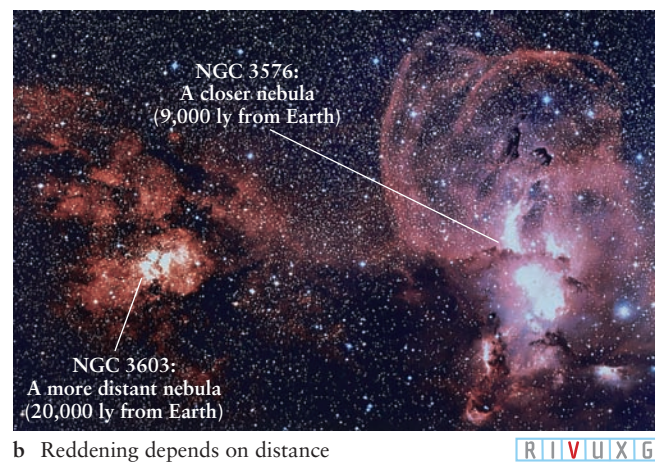


FIGURE 11-4 Interstellar Reddening (a) Dust in interstellar space scatters more short-wavelength (blue) light passing through it than longer-wavelength colors. Therefore, stars and other objects seen through interstellar clouds appear redder than they would otherwise. (b) Light



from these two nebulae passes through different amounts of interstellar dust and therefore they appear to have different colors. Because NGC 3603 is farther away, it appears a ruddier shade of red than does NGC 3576. (b: Australian Astronomical Observatory/David Malin Images)

Having located interstellar matter, we will now consider why some of this gas and dust becomes Jeans unstable and, therefore, collapses to form new stars (and planets). We explore this activity by expanding on the material in Section 4-2.

11-2 Supernovae, collisions of interstellar clouds, and starlight trigger new star formation

We will see in detail in Chapter 13 that a supernova is a violent detonation that ends the life cycle of a massive star. The core of the doomed star collapses in a matter of seconds, releasing vast quantities of particles and energy that blast the star's outer layers into space at speeds of several thousand kilometers per second.

1 Astronomers have found the ashes (more properly, the gas and dust) of many such dead stars scattered across the sky. These **supernova remnants** are another type of nebula. Supernova remnants, like the Cygnus Loop shown in **Figure 11-5**, have a distinctly arched appearance, as would be expected for an expanding shell of gas and dust. As it passes through the surrounding interstellar medium, the supernova remnant collides with preexisting matter, exciting the electrons in the atoms and molecules there, causing the gases to glow. If the expanding shell of a supernova remnant rams into a giant molecular cloud with sufficient speed, it can compress

that cloud, making pieces of it Jeans unstable and thus stimulating star birth in it. As we learned in Chapter 4, there is evidence that such an event happened around the time the solar system formed.

A simple collision between two interstellar clouds can also create regions that are Jeans unstable so that they collapse and form new stars. Likewise, radiation from O and B stars, which are especially bright and hot, will ionize (remove electrons from their atomic orbits) the gas that surrounds them, which then moves away and compresses the nearby interstellar medium (**Figure 11-6**).

2 As small regions of a giant molecular cloud become Jeans unstable and collapse, they become denser, preventing light from behind and inside them from escaping. These darker regions are called **Bok globules** (**Figure 11-7a**, see also **Figure 11-6**), named after astronomer Bart Bok, who first studied them in the 1940s. Infrared observations show compact regions of gas and dust, called *dense cores*, inside Bok globules. These cores are destined to become stars, as their gravity pulls their matter inward.

Often a giant molecular cloud has several hundred or even thousands of dense cores. In that case, hundreds or thousands of stars form together. Such stellar nurseries will become **open clusters** of stars like the Pleiades (see **Figure 11-2**). Open clusters are gravitationally unbound systems, meaning that the stars in them eventually drift

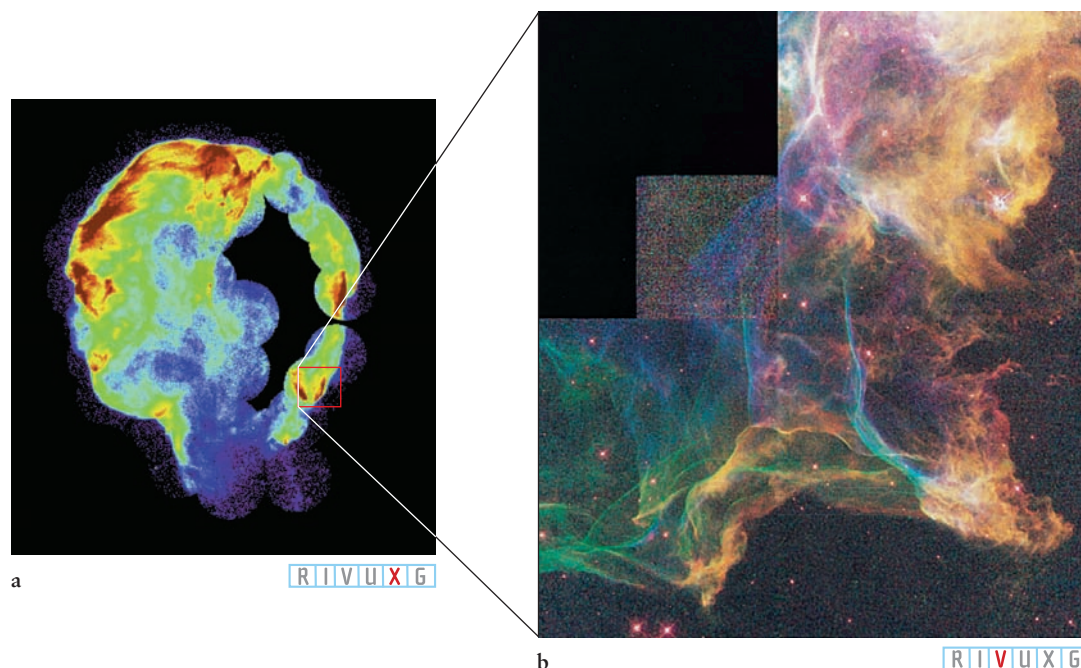


FIGURE 11-5 A Supernova Remnant (a) X-ray image of the Cygnus Loop, the remnant of a supernova that occurred nearly 20,000 years ago. The expanding spherical shell of gas now has a diameter of about 120 light-years. The entire Cygnus Loop has an angular diameter in our sky 6 times wider than the

Moon. (b) This visible-light Hubble Space Telescope image of part of the Cygnus Loop shows emission from different atoms false-color-coded with blue from oxygen, red from sulfur, and green from hydrogen. (a: Nancy Levenson/NASA; b: Jeff Hester [Arizona State University] and NASA)

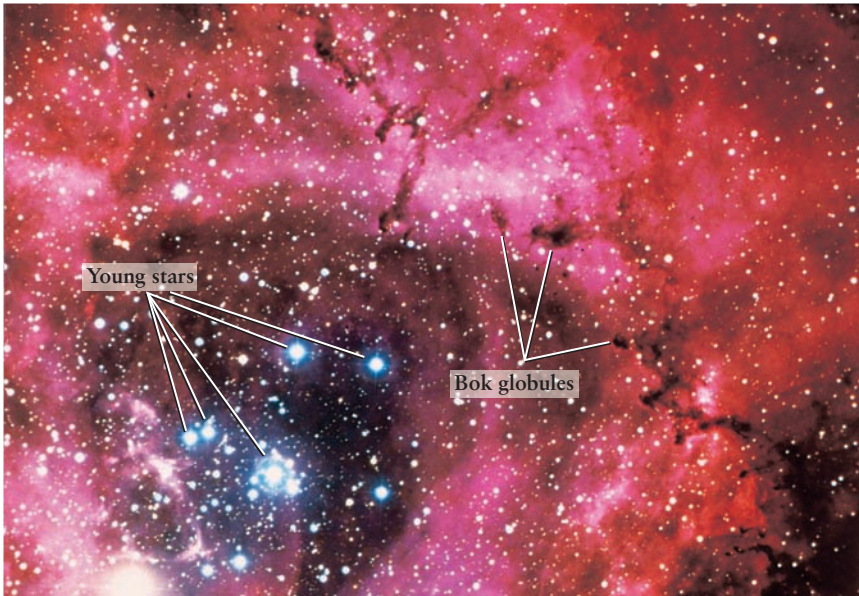


FIGURE 11-6 The Core of the Rosette Nebula The large, circular Rosette Nebula (NGC 2237) is near one end of a sprawling giant molecular cloud in the constellation Monoceros (the Unicorn). Radiation from young, hot stars has blown gas away from the center of this nebula. Some of this gas has become clumped in Bok globules that appear silhouetted against the glowing background gases. New star formation is taking place within these globules. The entire Rosette Nebula has an angular diameter on the sky nearly 3 times that of the Moon, and it lies some 3000 light-years from Earth. (Australian Astronomical Observatory/David Malin Images)

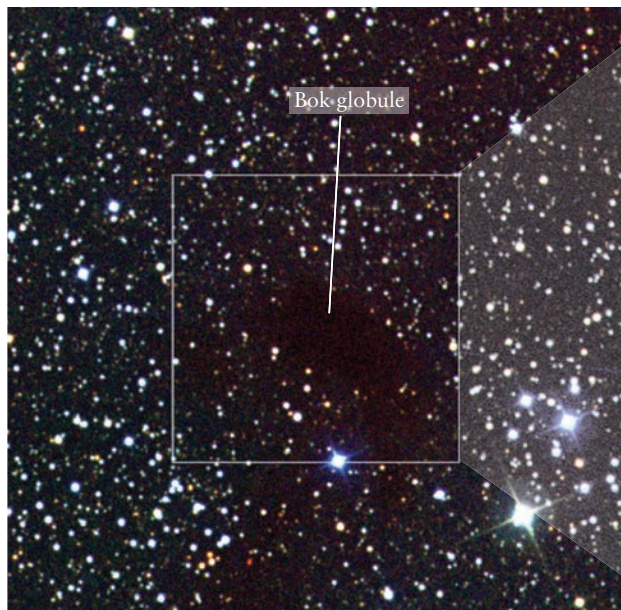
R I V U X G

apart. As noted in Chapter 4, it is likely that the solar system formed in an open cluster.

At first, a collapsing dense core is just a cool, dusty region thousands of times larger than our solar system. The dense core actually collapses from the inside out. The inner region falls in rapidly, leaving the outer layers of the dense core to drift in at a more leisurely rate. This process of increasing mass in the central region is called *accretion*, and the newly forming object at the center is called a **protostar** (Figure 11-7b). The formation of stars

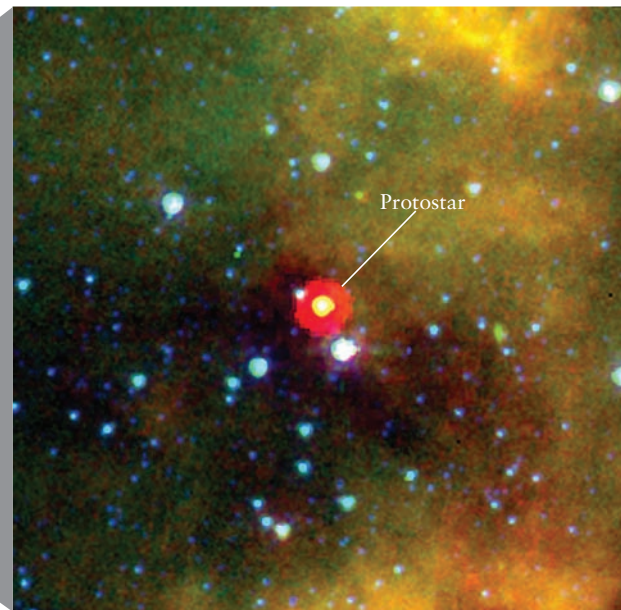
with different masses depends, in part, on the density of the interstellar medium from which the stars form. Denser gas in a Bok globule can lead to a more massive star forming there, provided that the globule has enough total mass.

Although fusion has not begun, a protostar emits energy, some of which comes from the compression and heating of its interior caused by the gravitational force from its growing mass of hot gas. However, most of the energy it releases comes from infalling gases colliding



a A dark nebula

R I V U X G



b A hidden protostar within the dark nebula

R I V U X G

FIGURE 11-7 Protostar in a Bok Globule (a) This visible-light image shows a small dark nebula (equivalently, a Bok globule) called L1014 located in the constellation Cygnus. (b) When

viewed in the infrared, a protostar is visible within the nebula. (a and b: NASA/JPL-Caltech/N. Evans [Univ. of Texas at Austin]/DSS)

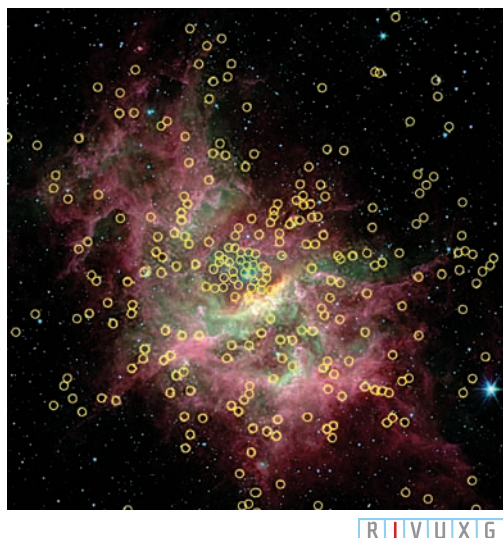


FIGURE 11-8 A Cluster of Protostars More than 300 protostars (yellow circles) were observed in the infrared by the Spitzer Space Telescope. This cluster of newly forming stars is 13,700 light-years away in the constellation Centaurus. The nebula, some of whose gas is being converted into stars, is called RCW 49 and contains more than 2200 stars and proto-stars. Most of the interior of this nebula is hidden from our eyes by the dust it contains. (NASA/JPL-Caltech/E. Churchwell, University of Wisconsin)

with the surface gas of the protostar. **Figure 11-8** is an infrared image showing the locations of myriad protostars in and around a nebula in Centaurus.

If a dense core is not spinning, it collapses into a sphere, which ultimately becomes an isolated star. If a dense core is spinning, it collapses into a disk, which may then condense into two or three stars. Or, if the disk has a low enough mass, it may become a single star with orbiting protoplanets, as we discussed in Chapter 4.

11-3 When a protostar ceases to accumulate mass, it becomes a pre-main-sequence star

Protostars are physically larger than the main-sequence stars into which they are evolving. A protostar of $1 M_{\odot}$, for example, has about 5 times the diameter of the Sun. Because of their large sizes, protostars emit great quantities of radiation and gas (analogous to solar wind), and they can be observed as sources of infrared radiation.

At this point, they cannot be seen in visible light because they are enshrouded by their outer layer of gas and dust (see Figure 11-7b). Much matter is still slowly falling inward from the dense core's outer shell. Eventually, the radiation and particles that flow off the protostar exert enough outward force to halt the infall of this gas and dust. As a

result, mass accretion stops, and the protostar becomes a **pre-main-sequence star**.

A pre-main-sequence star contracts slowly, unlike the rapid collapse of a protostar. When the temperature at its core reaches 10^7 K, hydrogen fusion begins. As we saw in Section 9-7, this thermonuclear process releases enormous amounts of energy. The outpouring of energy from hydrogen fusion creates enough pressure inside the pre-main-sequence star to stop its contraction. In the final stages of pre-main-sequence evolution, the outer shell of gas and dust finally dissipates (**Figure 11-9**; see also Figure 11-6). For the first time, the star is revealed via visible light to the outside universe.

11-4 The evolutionary track of a pre-main-sequence star depends on its mass

The more massive a pre-main-sequence star is, the more rapidly it begins hydrogen fusion in its core. For example, calculations indicate that a $5 M_{\odot}$ pre-main-sequence star starts fusing less than a million years after it first forms from a protostar, whereas a $1 M_{\odot}$ pre-main-sequence star takes a few tens of millions of years to begin fusion.

Calculations reveal that the minimum temperature required to start normal hydrogen fusion (see Discovery 9-1: Thermonuclear Fusion) in the core of a star is 10 million K. However, pre-main-sequence stars less massive than $0.08 M_{\odot}$ do not have enough gravitational

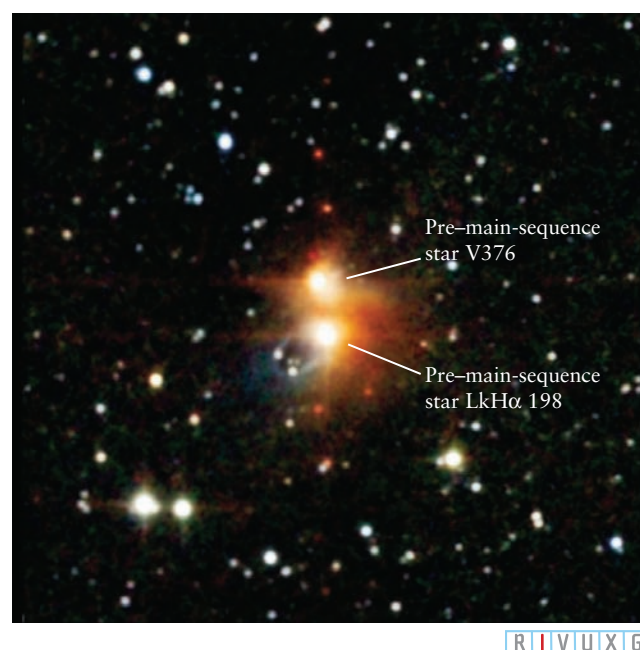


FIGURE 11-9 Pre-Main-Sequence Stars Seen in infrared, the two large bright objects in the center of this image are pre-main-sequence stars. They have recently shed their cocoons of gas and dust but still have strong stellar winds that create their irregular shapes. The two stars are an optical double; that is, they are not orbiting each other. (Atlas Image Courtesy of 2MASS/UMASS/IPAC-Caltech/NASA/NSF)

Focus Question 11-2

What types of events can initiate the process of star formation?

Focus Question 11-3

Why can't protostars be observed with visible-light telescopes?

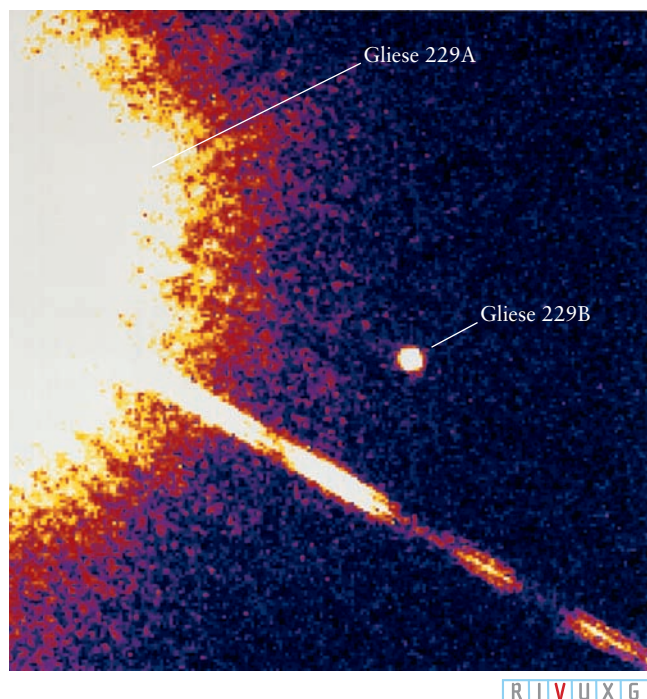


FIGURE 11-10 A Brown Dwarf Located 18 light-years (6 pc) from Earth in the constellation Lepus (the Hare), Gliese 229B was the first confirmed brown dwarf ever observed. With a surface temperature of about 1000 K, its spectrum is similar to that of Jupiter. Gliese 229B is in orbit around a star. The overexposed image of part of its companion, Gliese 229A, appears on the left. The two bodies are separated by about 43 AU. Gliese 229B has from 20 to 50 times the mass of Jupiter, but the brown dwarf is compressed to the same size as Jupiter. The spike of light was produced when Gliese 229A overloaded part of the Hubble Space Telescope's electronics. (S. Kulkarni [Caltech], D. Golimowski [JHU] and NASA)

force compressing and heating their cores to ever get this hot. As a result, these small bodies contract to become planetlike orbs of hydrogen and helium, called **brown dwarfs** (Figure 11-10). To date the lowest-mass star that has been observed has $0.091 M_{\odot}$. This star, part of a binary system, is only 95 times more massive and 16% larger than Jupiter.

Astrophysicists use computers and the equations of stellar structure (described in Section 9-8) to model the evolution of a pre-main-sequence star. By calculating changes in the energy that the contracting star emits, computer simulations can follow its changing position on a Hertzsprung-Russell (H-R) diagram (Figure 11-11). Keep in mind that such an **evolutionary track** represents changes in a star's temperature and luminosity, not its motion in space.

Protostars transform into pre-main-sequence stars as they cross a curve called the **birth line** (see the *blue line* in Figure 11-11). A star's exact location on this curve depends primarily on its mass and, to a much smaller extent, on the amount of metal it contains. (Recall from Section 10-6 that all elements other than hydrogen and helium are considered metals by astronomers.)

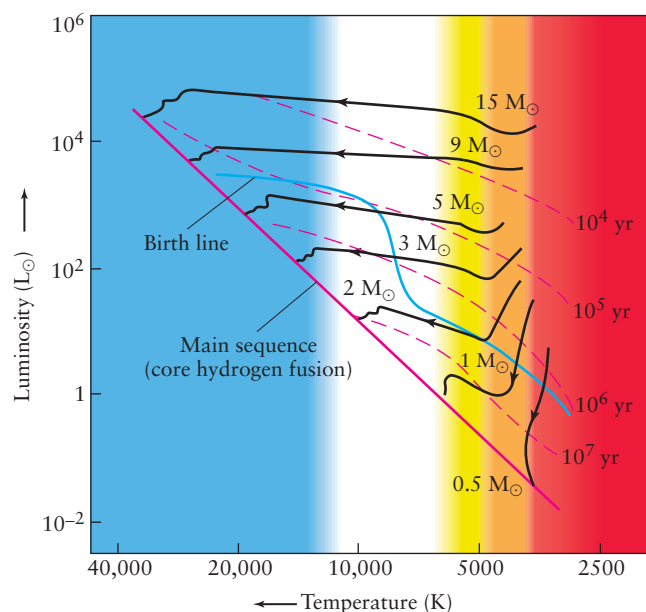


FIGURE 11-11 Pre-Main-Sequence Evolutionary Tracks This H-R diagram shows evolutionary tracks based on models of seven stars having different masses. The dashed lines indicate the stage reached after the indicated number of years of evolution. The birth line, shown in blue, is the location where each protostar stops accreting matter and becomes a pre-main-sequence star. Note that all tracks terminate on the main sequence at points that agree with the mass-luminosity relationship (see Figure 10-14a).

Spectroscopic observations of pre-main-sequence stars show many are vigorously ejecting gas, often in oppositely directed jets, just before they reach the main sequence. Gas-ejecting stars in spectral classes G and cooler (that is, G, K, and M) are called **T Tauri stars** (Figure 11-12a), after the first example discovered in the constellation of Taurus. Some astronomers propose that the onset of hydrogen fusion is preceded by vigorous chromospheric activity marked by enormous spicules (see Section 9-2) and flares (see Section 9-6) that propel the star's outermost layers back into space. If the protostar from which the T Tauri star condenses has a surrounding disk of gas and dust, as our Sun once had (see Sections 4-2 and 11-2), then much of the gas ejected back into space leaves as two jets perpendicular to the plane of the disk. These jets slam into surrounding interstellar gas, causing it to glow, thereby creating **Herbig-Haro objects** (see Figure 11-12b).

An infant star going through its T Tauri stage can lose as much as $0.4 M_{\odot}$ of matter and also shed its surrounding gas and dust while still a pre-main-sequence star. In light of all this activity, it is not surprising that observations reveal T Tauri stars to be variable, meaning that they change brightness much more than, say, the Sun does today. We will discuss variable stars further in Sections 11-12 and 11-13.

Focus Question 11-4

Which star arrives on the main sequence first, one that is $0.5 M_{\odot}$ or one that is $2 M_{\odot}$?

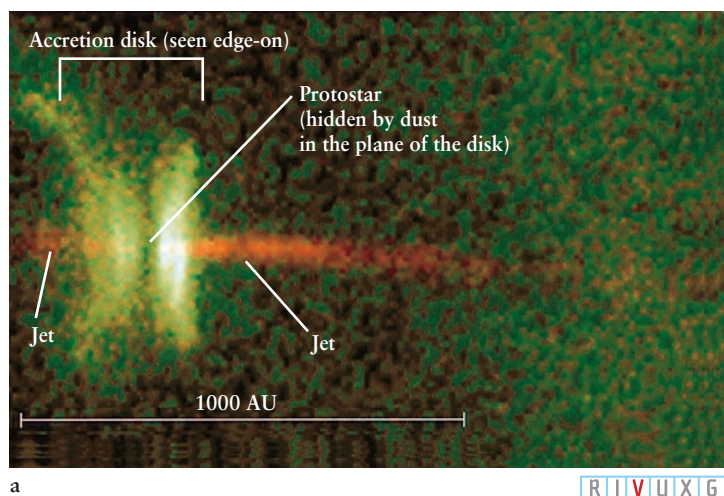


FIGURE 11-12 T Tauri Star (a) Protostar HH 30 is emitting gas primarily as two opposed jets. Called a T Tauri star, it is not yet on the main sequence. The star is hidden inside a disk of gas and dust. (b) The Herbig Haro objects HH1/HH2. These two glowing regions are on opposite sides of a T Tauri object in the constellation



Orion. The two outer edges of these objects are separated by about 1 light-year. (a: C. Burrows [STScI & ESA], the WFPC 2 Investigation Definition Team, and NASA b: J. Hester (Arizona State University), the WFPC 2 Investigation Definition Team, and NASA)

Focus Question 11-5

How are T Tauri stars different from main-sequence stars like the Sun?

Pre-main-sequence stars more massive than $2 M_{\odot}$ become hotter without much change in overall luminosity. The evolutionary tracks of these pre-main-sequence stars thus traverse the H-R diagram nearly horizontally, from right to left (see Figure 11-11). A star that is more massive than about $7 M_{\odot}$ has no pre-main-sequence phase at all. Its gravitational compression is so great that it begins to fuse hydrogen in its protostellar phase.

Until 2004, the upper limit to main-sequence stellar mass was thought to be around $120 M_{\odot}$. This number was based on calculations that show that above this mass, protostars rapidly develop extremely high fusion rates in their cores, which lead to extremely high surface temperatures—temperatures so great that their outer layers are superheated and thereby expelled into interstellar space. This expulsion, in turn, decreases their masses and their temperatures. An example of a very massive star in the process of shedding mass as it settles down onto the main sequence is the Pistol Star (Figure 11-13). One of the most luminous stars in our Galaxy, the Pistol Star may have started as a protostar with $200 M_{\odot}$. Observations reveal that every few thousand years it expels shells of gas, and it may have less than $10 M_{\odot}$ left when the expulsion of matter stops. The entire process of mass loss by very massive stars takes only a few million years.

In 2004, however, astronomers observed a main-sequence star that apparently has between 130 and $150 M_{\odot}$. In 2006, they saw the remnants of a star that must have been at least $150 M_{\odot}$. In 2010, the star R136a1 was observed to have $265 M_{\odot}$. As it has been shedding mass for some time, calculations indicate that

its original mass was $320 M_{\odot}$. Clearly, our understanding of star formation is very much a work in progress.

11-5 H II regions harbor young star clusters

We can detect young open clusters of stars from the magnificent glows they create in the nebulae in which

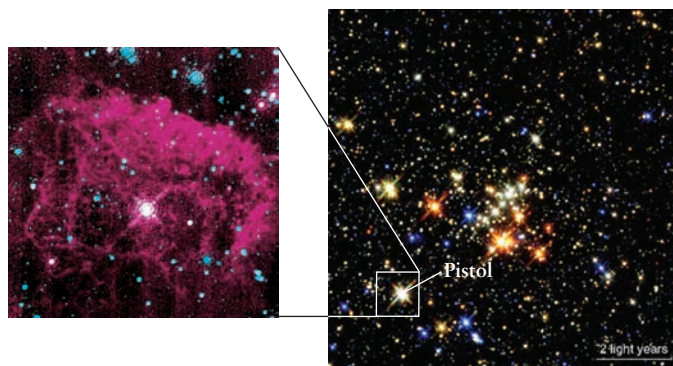


FIGURE 11-13 Mass Loss from a Supermassive Star The Quintuplet Cluster is 25,000 light-years from Earth. (Inset) Within the cluster is one of the brightest known stars, called the Pistol. Astronomers calculate that the Pistol formed nearly 3 million years ago and originally had 100 – $200 M_{\odot}$. The structure of the gas cloud suggests the star ejected the gas we see in two episodes, 6000 and 4000 years ago. The gas from any previous ejections is so thinly spread now that we cannot see it. The nebula shown in the inset is more than 4 ly (1.25 pc) across—it would stretch from the Sun nearly to the closest star, Proxima Centauri. The name Pistol was given to the star based on early, low-resolution radio images of its gas, which initially looked like an old-fashioned pistol aimed to the left near the top of the inset. (Inset: NASA; NASA and Don F. Figer [UCLA])



FIGURE 11-14 An H II Region This emission nebula, M16, called the Eagle Nebula because of its shape, surrounds a star cluster. It is so numbered because it was the sixteenth object in the Messier Catalogue of astronomical objects. Star formation is presently occurring in M16, which is located 7000 light-years from Earth in the constellation of Serpens Cauda (the Serpent's Tail). Several bright, hot O and B stars are responsible for the ionizing radiation that causes the gases to glow. (**Inset**)

Star formation is occurring inside these dark pillars of gas and dust. Intense ultraviolet radiation from existing massive stars off to the right of this image is evaporating the dense cores in the pillars, thereby prematurely terminating star formation there. Newly revealed stars are visible at the tips of the columns. (Australian Astronomical Observatory/David Malin Images; NASA, ESA, STScI, J. Hester and P. Scowen [Arizona State University])

they form. Figure 11-3 and **Figure 11-14** show examples of these emission nebulae. Because these nebulae are predominantly ionized hydrogen, they are also called **H II regions**. To see why H II regions occur, remember that the most massive pre-main-sequence stars, those of spectral types O and B, are exceptionally hot. Their surface temperatures are typically 15,000 to 35,000 K, causing them to emit vast quantities of ultraviolet radiation. This energetic radiation easily ionizes any surrounding hydrogen gas, thereby creating an H II region. Photons from an O5 star can ionize hydrogen atoms up to 30 light-years away.

H II denotes ionized hydrogen, which has no electrons in orbit that can make transitions and give off or absorb photons; therefore, how can we observe it? While some hydrogen atoms in the H II regions are being knocked apart by ultraviolet photons, some of the free protons and electrons approach each other so closely that they combine to become short-lived

neutral hydrogen, H I atoms (which are soon re-ionized by ultraviolet radiation). As these new hydrogen atoms assemble, their electrons return to their ground states ($n = 1$). This downward cascade through each atom's energy levels, releasing photons with each jump between levels, is what makes the nebula glow. Particularly prominent is the transition from $n = 3$ to $n = 2$, which produces H_α photons at 656 nm in the red portion of the visible spectrum (review the emission line spectrum in Figure 3-48c and d). Thus, the nebula around a newborn star cluster often shines with a distinctive reddish hue (see Figure 11-14). Nebulae that have oxygen often look green when seen with the eye through a telescope because this gas has a green emission line at 501 nm. Because the eye is more sensitive to green light than red, the dimmer oxygen emission line appears brighter in our brains than does the H_α line, which shows up well on CCD images, causing these lines to appear red.

An H II region is a small, bright “hot spot” in a giant molecular cloud. The collection of hot, bright O and B stars that produces the ionizing ultraviolet radiation is called an **OB association**. The famous Orion Nebula (**Figure 11-15**; see also Figure 1-4 and Figure 11-1) is an example. Four O and B stars in an open cluster called the Trapezium, at the heart of the Orion Nebula, are the primary sources of the ionizing radiation that causes the surrounding gases to glow. The Orion Nebula is embedded in a giant molecular cloud whose mass is estimated at $500,000 M_{\odot}$.

The OB association that creates an H II region also affects the rest of the giant molecular cloud in which it is embedded (see Figure 11-15 insets). Detailed models indicate that vigorous stellar winds, along with ionizing ultraviolet radiation from these stars, carve out a cavity in the cloud. Where this outflow is supersonic, it creates a shock wave, like the sonic boom created when a whip is snapped (for example, see the shock waves shown in the right inset of Figure 11-15). The shock wave forms along the outer edge of the expanding H II region, compressing hydrogen gas as it passes and thereby stimulating a new round of star birth. As more O and B stars form, they power the expansion of the H II region still farther into the giant molecular cloud. Meanwhile, the older O and B stars left behind begin to disperse (**Figure 11-16**). In this way, an OB association “eats into” a giant molecular cloud, creating stars in its wake. The inset in Figure 11-16 shows star formation around a single O star.

11-6 Plotting a star cluster on an H-R diagram reveals its age

As noted in Section 11-2, stars are often observed to form in open clusters, such as seen in the nearby Orion Nebula (see Figure 11-15). Numerous other star-forming regions have been identified, and the young open clusters in them offer astronomers a rich source of information about stars in their infancy. By measuring each star’s apparent magnitude, color, and distance, an astronomer can deduce its luminosity and surface temperature. The data

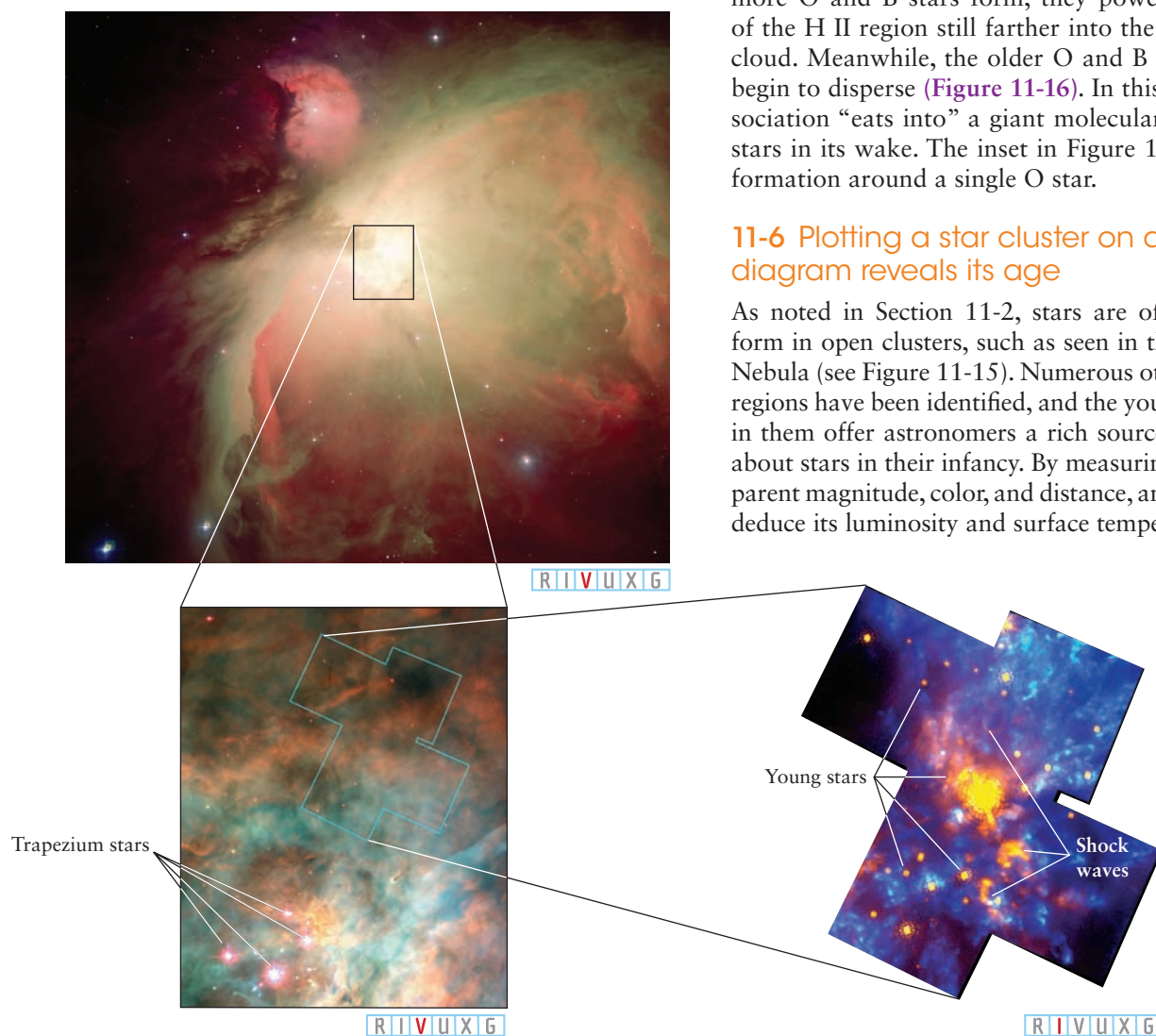


FIGURE 11-15 The Orion Nebula The middle “star” in Orion’s sword is actually the Orion Nebula, part of a huge system of interstellar gas and dust in which new stars are now forming. The Orion Nebula is a region visible to the naked eye. It is 1600 ly (490 pc) from Earth and has a diameter of roughly 16 ly (5 pc). This nebula’s mass is about $300 M_{\odot}$. (**Left inset**) This view at visible wavelengths shows the inner regions of the Orion Nebula. At the lower left are four massive stars, the brightest members of the Trapezium star cluster, which cause the nebula to glow. (**Right inset**) This view shows that infrared radiation penetrates

interstellar dust that absorbs visible photons. Numerous infrared objects, many of which are stars in the early stages of formation, can be seen, along with shock waves caused by matter flowing out of protostars faster than the speed of sound waves in the nebula. Shock waves from the Trapezium stars may have helped trigger the formation of the protostars in this view. (ESO; left inset: C. Robert O’Dell, Shui Kwan Wong [Rice University], and NASA; right inset: Rodger Thompson, Marcia Rieke, Glenn Schneider, Susan Stolovy [University of Arizona]; Edwin Erickson [SETI Institute/Ames Research Center]; David Axon [STScI]; and ESO)

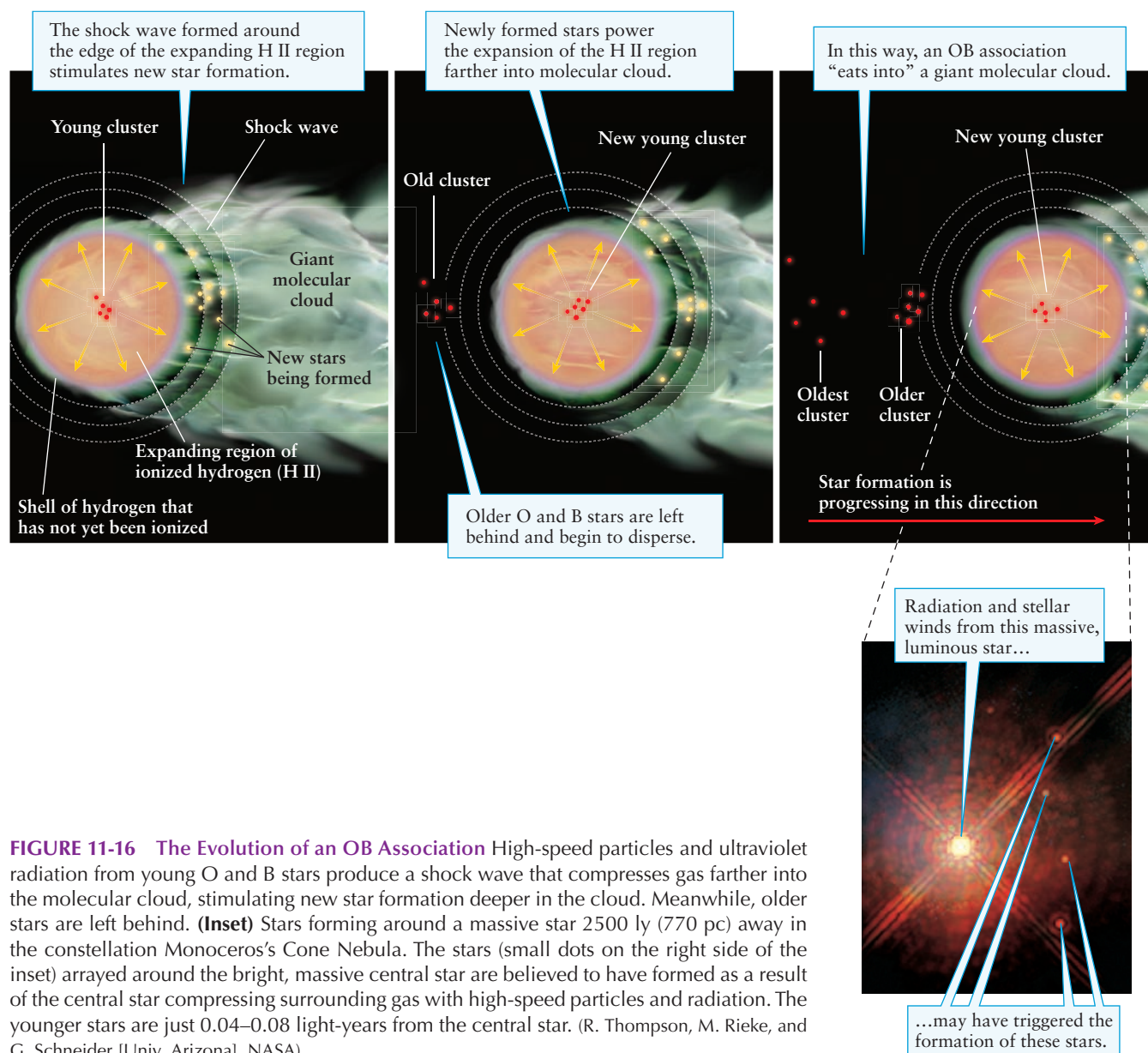


FIGURE 11-16 The Evolution of an OB Association High-speed particles and ultraviolet radiation from young O and B stars produce a shock wave that compresses gas farther into the molecular cloud, stimulating new star formation deeper in the cloud. Meanwhile, older stars are left behind. (*Inset*) Stars forming around a massive star 2500 ly (770 pc) away in the constellation Monoceros’s Cone Nebula. The stars (small dots on the right side of the inset) arrayed around the bright, massive central star are believed to have formed as a result of the central star compressing surrounding gas with high-speed particles and radiation. The younger stars are just 0.04–0.08 light-years from the central star. (R. Thompson, M. Rieke, and G. Schneider [Univ. Arizona], NASA)

for all the stars in the cluster can then be plotted on an H-R diagram, as shown for the Pleiades in Figure 11-2 or for the cluster NGC 2264 in Figure 11-17.

Because all of the stars in a cluster begin forming at the same time, and stars with different masses arrive on the main sequence at different times, astronomers can use the H-R diagram to determine the age of a cluster. For example, note that the hottest stars in NGC 2264 lie on the main sequence. These hot stars have surface temperatures around 20,000 K and are extremely bright and massive. Their radiation also causes the surrounding gases to glow. Most of the stars cooler than about 10,000 K have not yet arrived at the main sequence. These less massive stars, which are in the final stages of pre-main-sequence contraction, are just now beginning to ignite thermonuclear reactions at their centers.

From the H-R diagram of a young cluster we can see which are the lowest-mass stars that have already entered the main sequence. Using this information, along with theories of stellar evolution, we can determine the cluster’s age. The cluster NGC 2264, for example, is roughly 2 million years old. In contrast, nearly all of the stars in the Pleiades (see Figure 11-2b) have completed their pre-main-sequence stage. That cluster’s age is calculated to be about 100 million years, which is how long it takes for the least massive stars to finally begin hydrogen fusion in their cores.

Open clusters dissipate. As noted earlier, open clusters, such as the Pleiades and NGC 2264, possess barely enough mass to hold themselves together. A star moving

Focus Question 11-6

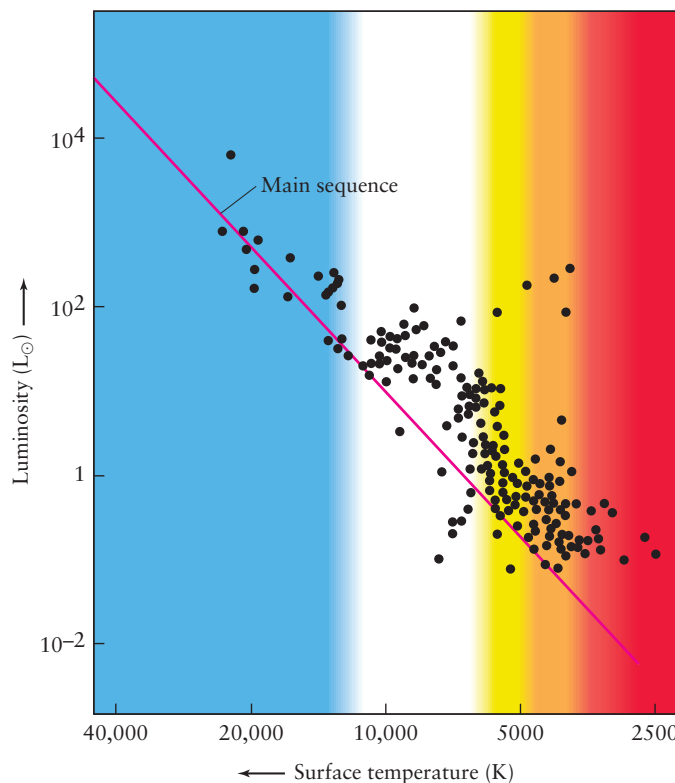
What are groups of high-mass stars called?



a NGC 2264

R I V U X G

FIGURE 11-17 Plotting the Ages of Stars (a) This photograph shows a region of ionized hydrogen and the young star cluster NGC 2264 in the constellation Monoceros. The red nebula is located about 2600 light-years from Earth and contains numerous stars that are about to begin hydrogen fusion in their cores. (b) Each dot plotted on this H-R diagram



b

represents a star in this cluster whose luminosity and surface temperature have been measured. Note that most of the cool, low-mass stars have not yet arrived at the main sequence. Calculations of stellar evolution indicate that this star cluster started forming about 2 million years ago. (a: Australian Astronomical Observatory/David Malin Images)

faster than the average speed for the cluster occasionally escapes. This lowers the total gravitational force of the cluster, making it easier for other stars to leave. Observations indicate that the stars in most open clusters separate from each other and eventually mix with the rest of the stars in the Galaxy within 10–50 million years after the cluster began forming, although some open clusters last for a few hundred million years.

MAIN-SEQUENCE AND GIANT STARS

In the first part of this chapter, we saw how stars form from dense cores of gas and dust inside giant molecular clouds. As these cores collapse, the gas and dust in their centers quickly form protostars. Eventually, protostars have pulled in (accreted) most of the mass available to them and they become pre-main-sequence stars that slowly contract. As a pre-main-sequence star evolves, fusion begins in its core.

Recall from Chapter 9 that in our Sun, thermal pressure created by fusion in the core pushes outward, everywhere balancing the inward force of gravity so that the Sun neither expands nor collapses. The Sun is

said to be in hydrostatic equilibrium (see Figure 9-21). When each layer of a pre-main-sequence star can finally support all of the layers above it (meaning that collapse ceases), that star also comes into hydrostatic equilibrium, and a main-sequence star is born (see Figure 11-11). *Main-sequence stars are those stars in hydrostatic equilibrium, with nuclear reactions fusing hydrogen into helium in their cores at nearly constant rates.* Figure 11-18 summarizes the star formation process.

11-7 Stars spend most of their lives on the main sequence

The **zero-age main sequence (ZAMS)** is the set of locations on the H-R diagram where pre-main-sequence stars of different masses first become stable objects, neither shrinking nor expanding. This sequence is denoted by the solid red line that appears on several of the H-R diagrams in this chapter. Note that the theoretical evolutionary tracks in Figure 11-11 end at locations along the main sequence that agree with the observed mass-luminosity relation (recall Figure 10-14a): The most massive main-sequence stars are the most luminous, whereas the least massive stars are the least luminous.

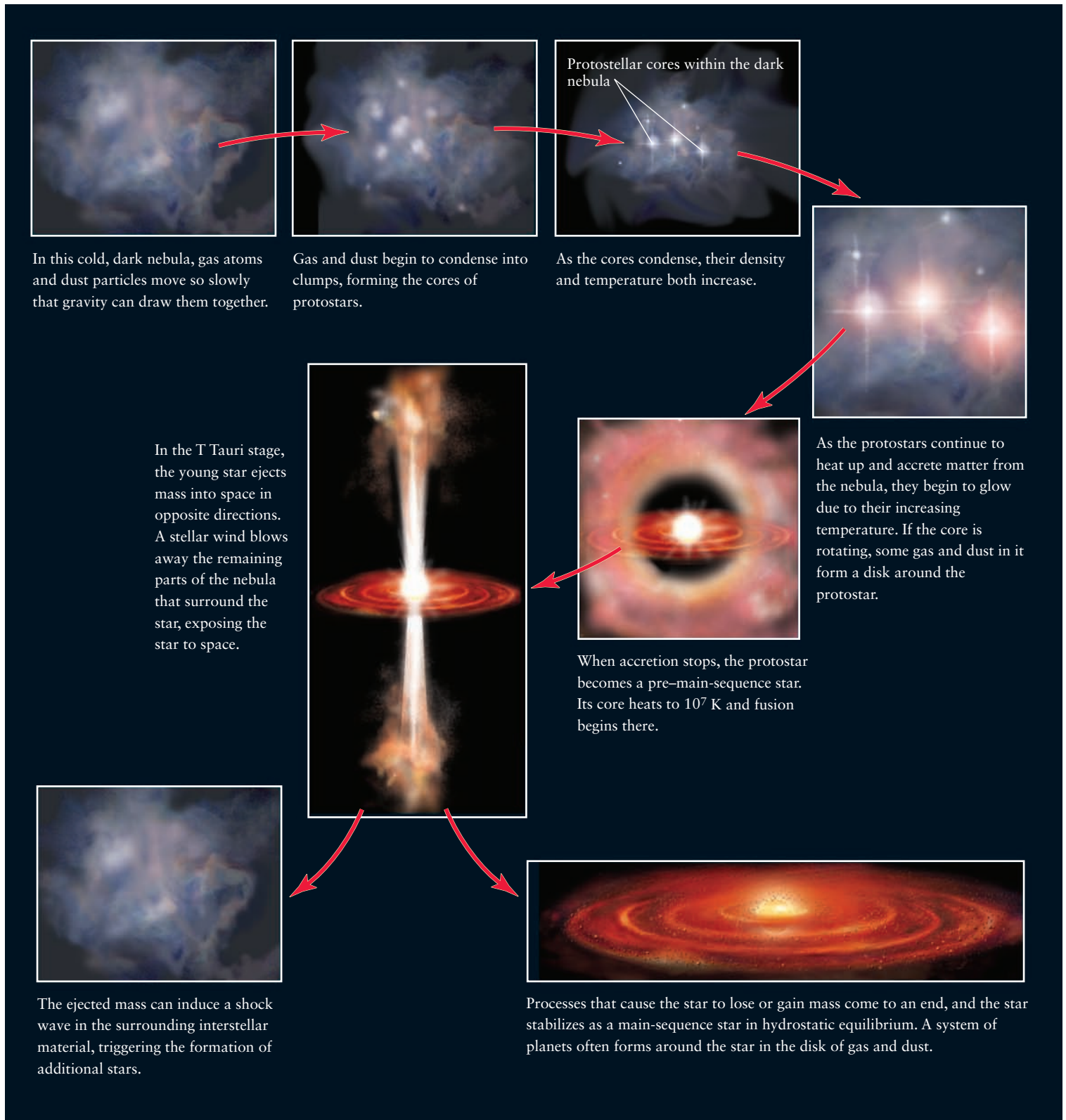


FIGURE 11-18 A Summary of the Star Formation Process This set of drawings takes you through the star formation process for the Sun and other stars with less than about 1.5 solar masses.

3 The energy emitted by stars is primarily generated by thermonuclear fusion (see Discovery 9-1: Thermonuclear Fusion). More luminous (brighter) stars generate and radiate more energy each second than do dimmer stars.

Put another way, higher-mass stars are consuming more fuel than are lower-mass stars. This raises the interesting question of whether higher-mass stars have enough extra nuclear fuel to shine longer than lower-mass stars.

TABLE 11-2 MAIN-SEQUENCE LIFETIMES

Mass ($M_{\odot}$)	Surface temperature (K)	Luminosity ($L_{\odot}$)	Time on main sequence (10^6 years)	Spectral class
25	35,000	80,000	3	O
15	30,000	10,000	15	B
3	11,000	60	500	A
1.5	7000	5	3000	F
1.0 (Sun)	6000	1	10,000	G
0.75	5000	0.5	15,000	K
0.50	4000	0.03	200,000	M

The answer is *no*. The equations of nuclear fusion reveal that higher-mass stars consume fuel so quickly that these stars actually spend *less* time at each stage of stellar evolution than do lower-mass stars. For example, the tremendous internal pressure created by gravity inside

O stars forces them to consume all of their core hydrogen in only a few million years, as shown in Table 11-2. Conversely, stars of very low mass take hundreds of billions of years to convert their cores from hydrogen into helium.

Because, as we will see, the universe is roughly 14 billion years old, no main-sequence star with mass less than about $0.75 M_{\odot}$ has yet moved into the next stage of stellar evolution.

The equations further reveal that the conversion of hydrogen into helium in every star’s core takes a long time compared to any other stage of its stellar evolution. (These stages are presented in the following sections and the next chapter.) That is why the vast majority of stars represented on an H-R diagram are located on the main sequence. This raises another intriguing question: Because most stars are on the main sequence, why are most of the stars visible to the naked eye non-main-sequence stars? The answer will become apparent shortly.

EVOLUTION OF LOW MASS
($0.08\text{--}0.4 M_{\odot}$) STARS

Because stars with mass less than $0.4 M_{\odot}$ and stars with mass more than $0.4 M_{\odot}$ evolve beyond the ZAMS so differently, we explore their evolutionary stages separately.

11-8 Red dwarfs convert essentially their entire mass into helium

The lowest-mass main-sequence stars, called **red dwarfs**, have masses between $0.08 M_{\odot}$ and $0.4 M_{\odot}$. They have the lowest core temperatures of all stars. Because fusion rates depend critically on temperature, these stars produce the

least amount of energy and therefore emit the least energy; they are the dimmest of all main-sequence stars.

Red dwarfs are also different from higher-mass main-sequence stars in that the lower-mass stars transport the helium that they create out of their cores. This hot helium moves upward by convection (see Section 6-4), while cooler, denser gases from the hydrogen-rich outer layers move downward and into the core (Figure 11-19). This hydrogen is also fused into helium. As a result, these stars eventually convert their entire mass into helium. In other words, they are fully convective.

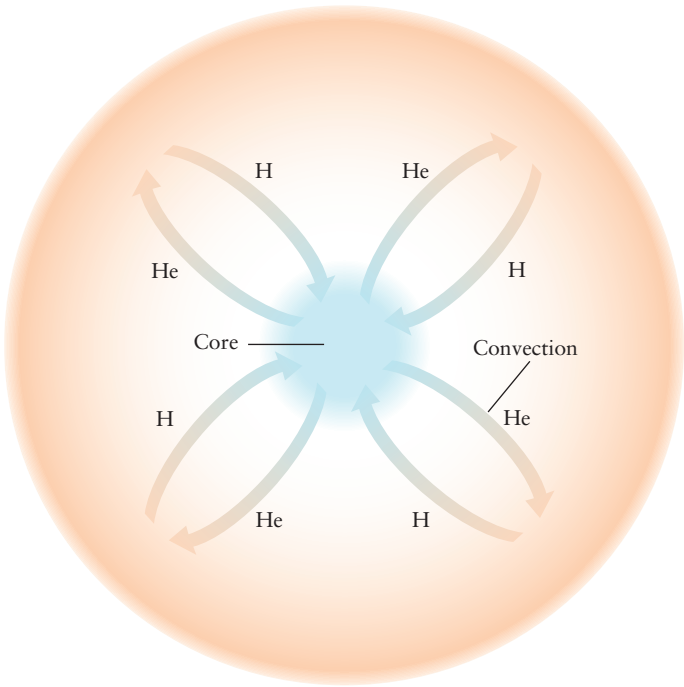


FIGURE 11-19 Fully Convective Star This drawing shows how the helium created in the cores of red dwarfs rises into the outer layers of the star by convection, while the hydrogen from the outer layers descends into the core. This process continues until the entire star is composed of helium.

Fusion in the cores of red dwarfs is so slow that they remain on the main sequence for trillions of years. (In comparison, the theory of solar evolution predicts that the Sun's total lifetime on the main sequence will be about 10 billion years.) When a red dwarf has eventually converted its entire mass into helium, fusion will cease inside it because the star does not have enough pressure in its core to heat and fuse that helium into anything else. These helium bodies will then radiate the heat they generated while on the main sequence, becoming cooler and dimmer. They will move down and to the right on the H-R diagram.

We saw in Section 10-11 how there is a relationship between the masses and luminosities of main-sequence stars (for example, see Figure 10-14). Besides being the lowest-mass, dimmest main-sequence stars, red dwarfs are also the most common stars in our Galaxy, representing some 85% of all its stars. However, red dwarfs are so dim that astronomers continue to discover more of them even in our astronomical neighborhood. Red dwarfs typically form as isolated bodies surrounded with planets, unlike more massive stars, which are more apt to form in binary star systems. Because red dwarfs are the most common stars and are usually single, astronomers have recently concluded that most stars in our Galaxy are isolated, rather than binary star systems, as they thought until just a few years ago. Current calculations indicate that two-thirds of the stars in our Galaxy are isolated.

EARLY AND MIDDLE EVOLUTION OF INTERMEDIATE ($0.4\text{--}8\text{ }M_{\odot}$) AND HIGH-MASS STARS

The changes in stars with more than $0.4\text{ }M_{\odot}$ are much more complicated. In this chapter, we consider what happens to these stars on the main sequence and in the evolutionary stage after it. In the following chapter, we explore the ends of their lives.

11-9 When core hydrogen fusion slows down, a main-sequence star with mass greater than $0.4\text{ }M_{\odot}$ becomes a giant

Recall that the main sequence is defined by the fusion of hydrogen into helium in a star's core. Unlike lower-mass stars, stars with more than $0.4\text{ }M_{\odot}$ do not transport helium out of their cores by convection. Therefore, the chemical composition of their cores differs from the composition of their outer layers throughout their lives.

When the hydrogen in the core of a main-sequence star with more than $0.4\text{ }M_{\odot}$ is mostly converted into helium, fusion in the core drastically slows down. The temperature of the core at that time is not high enough to enable its helium to fuse into other elements. However, the star continues to evolve and, contrary to intuition, the end of core hydrogen fusion actually causes these stars to expand in size.

Recall from our discussion of the Sun in Chapter 9 that, while on the main sequence, a star's outer layers are supported by thermal pressure, the energy for which is supplied by fusion in the core. As the rate of this fusion decreases, a star can no longer support the crushing weight of its outer layers, and so the hydrogen-rich gas just outside the core is compressed inward and thereby heated enough to begin fusing into helium (Figure 11-20). This process is called **hydrogen shell fusion** because it occurs in a shell a few thousand kilometers thick surrounding the core. Recall from Discovery 9-1: Thermonuclear Fusion that the proton-proton chain is the major source of energy inside the Sun and all other stars fusing H to He in their cores. This fusion process also predominates in hydrogen shell fusion.

4 Why, then, does the star expand to become a giant? The photons created in shell fusion are generated closer to the star's outer layers than are the photons generated in the core. As noted in Section 9-8, photons provide the energy and hence the outward pressure to keep stars in hydrostatic equilibrium. Photons first created in the shell have more energy with which to push the outer layers outward than do photons from the core, which, created closer to the star's center, have to hold out more of the star's mass. In other words, the core photons have lost energy by the time they emerge from the core, so they cannot push the outer layers out as much as can shell photons. The outer layers of stars

Focus Question 11-8

Is the name *red dwarf* appropriate for the stars so named? Why or why not?

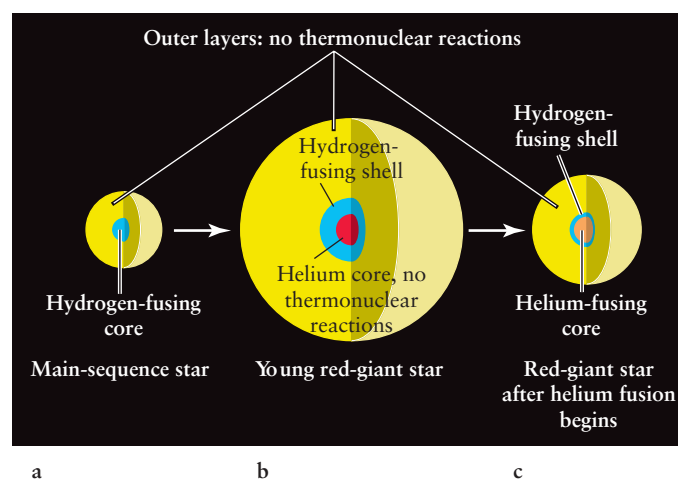


FIGURE 11-20 Evolution of Stars Off the Main Sequence

(a) Hydrogen fusion occurs in the cores of main-sequence stars. (b) When the core is converted into helium, fusion there ceases while the same fusion begins in a shell just outside the core. The star expands into the giant phase. This newly formed helium sinks into the core, which heats up. (c) Eventually, the core reaches 10^8 K , and core helium fusion begins. This activity causes the core to expand, slowing the hydrogen shell fusion and thereby forcing the outer layers of the star to contract.

with shell fusion therefore expand farther out (the stars become larger than they were on the main sequence).

Far from the fusing shell, the surface gases cool and the temperature of the star's bloated surface decreases between 3000 and 6000 K, depending on the star's total mass. In about 5 billion years, our Sun will have a helium core and will swell into a giant with a radius of about $\frac{1}{2}$ AU, vaporizing Mercury and perhaps causing Venus to spiral into the Sun. Earth will be scorched to a cinder.

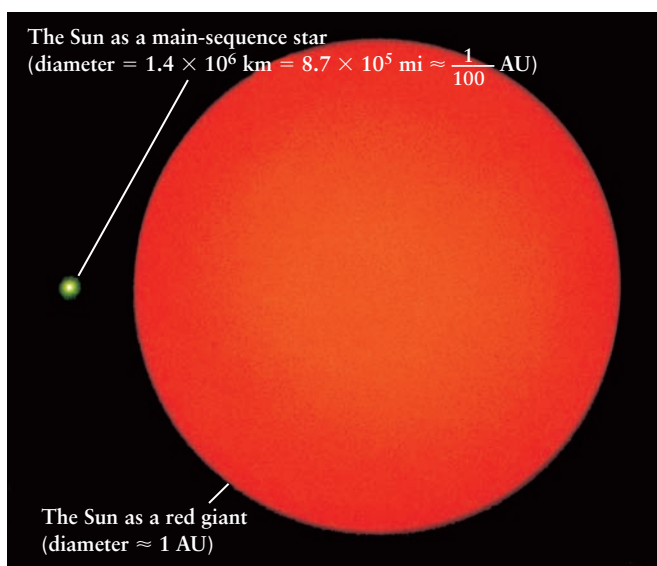
Giant stars are so enormous that their outer layers constantly leak gases into space. At times, this mass loss is significant (Figure 11-21). When studied spectroscopically, the escaping gases exhibit narrow absorption lines, and the lines from gases coming toward us are slightly blueshifted, due to the Doppler effect. This small shift corresponds to a speed of 10 km/s, typical of the expansion velocities with which gases leave the tenuous outer layers of giants. A typical rate of mass loss for a giant is roughly $10^{-7} M_{\odot}$ per year. For comparison, in a main-sequence star, such as the Sun, mass loss rates are only around $10^{-14} M_{\odot}$ per year.

Although the surface of a giant is cooler than that of the main-sequence star from which it evolved, the giant is more luminous: It can emit more photons

each second because it has so much more surface area. As a full-fledged giant (Figure 11-22a), our Sun will shine 2000 times more brightly than it does today.

Focus Question 11-9

Where in the Sun will the energy be generated to expand it into the giant phase?



a The Sun today and as a red giant

FIGURE 11-22 The Sun Today and as a Giant (a) In about 5 billion years, when the Sun expands to become a giant, its diameter will increase a hundred-fold from what it is now, while its core becomes more compact. Today, the Sun's energy is produced in a hydrogen-fusing core whose diameter is about 200,000 km. When the Sun becomes a giant, it will draw its energy from a



FIGURE 11-21 A Mass-Loss Star A red giant star is shedding its outer layers, thereby creating this reflection nebula, labeled IC 2220 and called Toby Jug, located in the constellation Carina. The star is embedded inside the nebula and is not visible in this image. (Australian Astronomical Observatory/David Malin Images)

11-10 Helium fusion begins at the center of a giant

When a star first becomes a giant, its hydrogen-fusing shell surrounds a small, compact core a few times the diameter of Earth, composed of almost pure helium. At



b Red giant star in the star cluster M103

hydrogen-fusing shell that surrounds a compact helium-rich core. The helium core will have a diameter of only 30,000 km. The Sun's diameter will be about 100 times larger, and it will be about 2000 times more luminous as a giant than it is today. **(b)** This image shows a red giant star in the cluster M103 in the constellation of Cassiopea. (b: NOAO/Science Source)

first, no thermonuclear reactions occur in the helium-rich core of a giant because the temperature there is too low to fuse helium nuclei. The hydrogen-fusing shell creates more helium, which settles into the core. The eventual transformation of the core follows two dramatically different routes, depending on the star's total mass. The lower-mass giants, from 0.4 to about 2 solar masses, go down an exceptionally bizarre path.

0.4 $M_{\odot}$ to 2 $M_{\odot}$ —helium flash The helium atoms in cores of young giant stars with masses between 0.4 and about 2 $M_{\odot}$ are squeezed by the gravitational force into a crystal-like solid. At the pressures they are under, these atoms are completely ionized, separating into nuclei and electrons. The electrons, distributed between the nuclei, are so closely crowded together that they come under the influence of a phenomenon called the **Pauli exclusion principle**, first formulated in 1925 by the Austrian physicist Wolfgang Pauli. According to this principle, nature does not allow two identical particles to exist in the same place at the same time. As the electrons are pressed closer and closer together, the exclusion principle forces many of them to move faster and faster so that they do not become “identical,” here meaning being in the same place and moving with the same speeds as adjacent electrons. This motion creates increased repulsion between adjacent electrons.

Astronomers call the electrons in this situation *degenerate*. Degeneracy, by keeping electrons apart, provides a pressure in the core and thereby prevents the core from collapsing. Thus, the helium-rich core of a low-mass giant is supported by **electron degeneracy pressure**. This state provides a greater outward pressure than did the normal pressure in the star before degeneracy developed in its core. The solid core grows from infalling helium and heats up under the gravitational influence of the star's mass.

Here is the weirdest part. The equations predict that, unlike the pressure of an ordinary gas, the pressure inside a degenerate core does not change with temperature. In a normal gas, like the air we breathe, when the temperature increases, the pressure also increases, causing the gas to expand and cool. When a degenerate core's temperature increases, the pressure there does not increase. Without the “safety valve” of increasing pressure, the star's core cannot expand and cool when overheated. Rather it sits there, just getting hotter and hotter.

Eventually, the core temperature reaches about 100 million K, at which time **core helium fusion** begins (see Figure 11-20c). This fusion creates energy, so the core's temperature increases. But a degenerate core's pressure is unable to increase. If the core pressure increased, the core would expand and cool. Because the core does not immediately expand, during the first few hours of core helium fusion, both the core temperature and the fusion rate increase dramatically. This situation is called the **helium flash**.

When the core temperature reaches around 3.5×10^8 K, the degenerate helium is forced to transform back into an ordinary gas. Suddenly, the usual safety valve operates, namely, that the high temperature increases the core pressure, causing the core to expand and cool, moderating the rate of fusion occurring there within a few hours. The expanding core, in turn, enlarges and cools the hydrogen-fusing shell around it. Fusion in the shell therefore decreases, so the flow of photons from the shell directly to the outer layers of the star drops, causing the outer regions to cool. As they cool, they contract—the star shrinks until it reaches a new hydrostatic equilibrium.

Keep in mind that the helium flash occurs in and acts upon the star's core. It does not cause a sudden flash in brightness. A 0.4 to 2 $M_{\odot}$ helium-fusing giant is left smaller, dimmer, and hotter than it was before it began fusing helium. **Figure 11-23a** shows the expected changing brightness of the Sun when going through this process. **Figure 11-23b** includes evolutionary tracks of two stars in this mass range as they leave the main sequence, move into the giant phase, go through the helium flash, and begin to settle back down. We will follow their evolution through the rest of the giant phase after briefly exploring how stars with more than 2 $M_{\odot}$ proceed off the main sequence.

More than 2 $M_{\odot}$ —no helium flash The pressure created by the gravitational force of a star with more than 2 solar masses is enough to smoothly compress and heat its core until it reaches the fusion temperature of helium. In other words, the cores of giants in this mass range never become degenerate.

When an ordinary gas, such as the helium that exists in the cores of these stars, is compressed, it heats up; conversely, when it expands, it cools down. If the energy production created by helium fusion overheats its core, that extra heat provides extra pressure that forces the core to expand, thereby cooling its gases and slowing the rate of thermonuclear reactions. Conversely, if too little energy is being created to support the star's overlying layers, the star's gravity pulls them farther inward, thereby compressing the core. The resulting increase in core temperature speeds up the thermonuclear reactions and thus increases the energy output, which stops the contraction. Either way, the “safety valve” keeps it from collapsing or exploding.

Figure 11-23b shows the evolution off the main sequence of three stars in this mass range, which do not undergo helium flash. These stars move up and to the right across the giant region of the H-R diagram, getting larger until core fusion begins, whereupon they contract slightly—less than stars that undergo the helium flash.

11-11 Life in the giant phase has its ups and downs

The dominant process for fusing helium in the cores of giants is called the *triple alpha process*, in which three

Focus Question 11-10

Will the Sun undergo a helium flash? Why or why not?

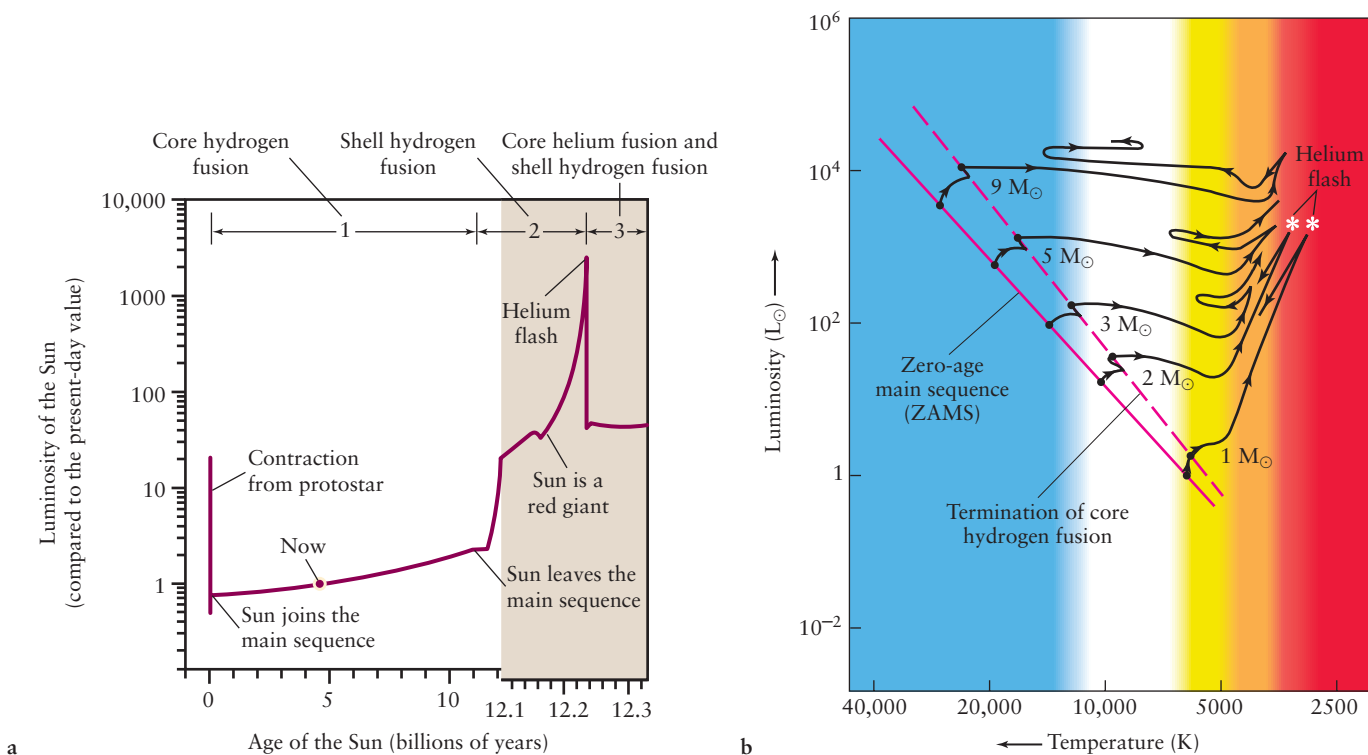
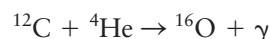


FIGURE 11-23 Post-Main-Sequence Evolution (a) The luminosity of the Sun changes as our star evolves. It began as a protostar with decreasing luminosity. On the main sequence today, it gradually brightens. Giant-phase evolution occurs more rapidly, with faster and larger changes of luminosity. Note

helium atoms fuse to become a carbon atom. (Early in the twentieth century, certain particles emitted by radioactive nuclei were called *alpha particles*. They were later identified as helium nuclei.) The triple alpha process works in two steps. First, two helium nuclei fuse to create beryllium. If within 10^{-8} s a third helium nucleus collides with the short-lived beryllium, then carbon is formed. Otherwise the beryllium decays back into two helium atoms. This process also releases energy and can be summarized as follows:



where γ (gamma) denotes energy emitted as photons. The fusion in giants does not stop there. Some of the carbon thus created can fuse with another helium nucleus to produce oxygen:



As in main-sequence stars, the energy released in these fusion processes is in the form of gamma rays.

Giant stars fuse helium in their cores for about 10% as long as they spend fusing hydrogen while on the main sequence. For example, the Sun will spend a total of about 10 billion years on the main sequence, after which

the change in scale of the horizontal axis at 12 billion years. (b) Model-based evolutionary tracks of five stars are shown on this H-R diagram. In the high-mass stars, core helium fusion ignites smoothly where the evolutionary tracks make a sharp turn upward into the giant region of the diagram.

it will be a giant for about 1 billion years. While helium fusion is occurring in a giant's core, further gravitational contraction of the star's core ceases.

Despite the fact that all stars spend most of their lives on the main sequence, most of the stars visible to the naked eye are giants! This occurs because the relatively few giants in our neighborhood are so luminous that they outshine their neighbors in the night sky.

VARIABLE STARS

After core helium fusion begins, the evolutionary tracks of mature stars move partway back toward the main sequence (right to left on an H-R diagram; Figure 11-23b). Their pressure-temperature safety valves are not perfect, so as they shrink, they overheat and core fusion causes them to expand to larger sizes than they would be at equilibrium. Becoming too large, their pressure and temperature decrease and they collapse, overcompressing and thereby overheating themselves again. This process is analogous to dropping a tennis ball, which bounces repeatedly after striking the ground. In other words, a star under these conditions becomes unstable and pulsates. The region on the H-R diagram between the main sequence and the right side of the giant branch, where this occurs, is called the **instability strip** (Figure 11-24).

Focus Question 11-11

How many helium atoms does it take to make one oxygen atom?

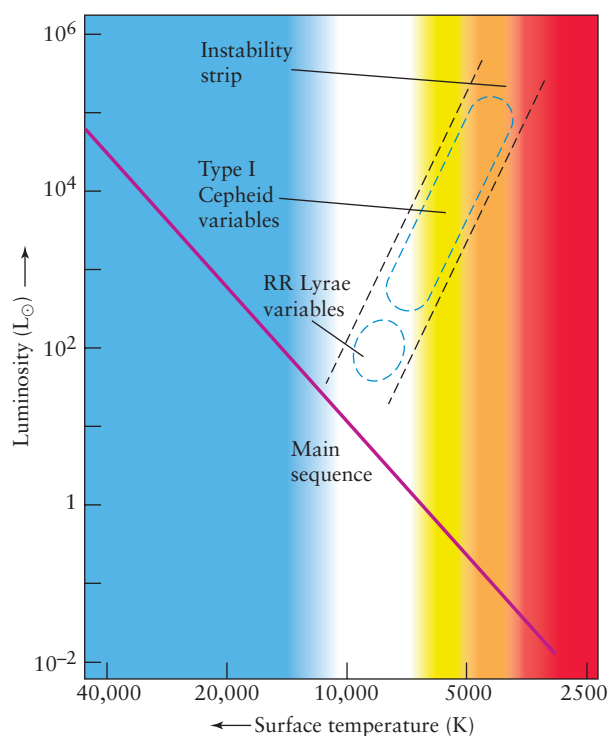


FIGURE 11-24 The Instability Strip The instability strip occupies a region between the main sequence and the giant branch on the H-R diagram. A star passing through this region along its evolutionary track becomes unstable and pulsates.

These so-called **variable stars** can be easily identified by their changes in brightness amid a field of stars of constant luminosity.

Lower-mass, post-helium-flash stars pass through the lower end of the instability strip as they move horizontally on the H-R diagram along their evolutionary tracks. These stars become **RR Lyrae variables**, named after the prototype in the constellation of Lyra (the Lyre). RR Lyrae variables all have periods shorter than 1 day. Higher-mass stars that evolve from the same interstellar medium as RR Lyrae pass back and forth through the upper end of the instability strip on the H-R diagram. These stars become **Type I Cepheid variables**, often simply called **Cepheids**. (There is another whole group of variable stars called Type II Cepheids that formed earlier in the history of the universe from gases with virtually no metals. We discuss them in more detail shortly.)

11-12 A Cepheid pulsates because it is alternately expanding and contracting

A **Cepheid variable** is characterized by the way in which its light output varies—most of these stars follow a pattern of rapid brightening followed by gradual dimming. A Cepheid variable brightens and fades because the star's outer layers cyclically expand and contract. Lines in the spectrum of δ (delta) Cephei shift back and forth with the same 5.4-day period that characterizes its variations

in magnitude. According to the Doppler effect measured from this star, these shifts mean that the star's surface is alternately approaching and receding from us. When a Cepheid variable pulsates, its gases alternately heat up and cool down; the surface temperature cycles between about 6300 K and 5000 K. Thus, the characteristic light curve (luminosity versus time curve, as in Figure 10-11) of a Cepheid variable results from changes in both size and surface temperature.

Just as a bouncing ball eventually comes to rest, a pulsating star would soon stop pulsating without some mechanism to maintain its oscillations. Variable stars have a layer of gas rich in partially ionized helium (helium with one electron stripped off). This gas, absorbing a lot of photons leaving the core, heats, expands, and pushes the outer layers outward, just as steam raises the lid of a pot. Eventually, the ionized helium layer is spread so thin that photons pass through it. The ionized helium cools and contracts, and the star's outer layers collapse, compressing the interior gases until the process repeats itself. In our analogy with the pot of boiling water, when enough of the steam inside escapes, the lid falls back and the steam inside builds up again. Variability only occurs when the conditions of temperature and pressure inside the star are appropriate. As a result, giants are variable only when the stars are in the instability strip of an H-R diagram.

Focus Question 11-12

How do astronomers observe that Cepheids are changing size?

11-13 Cepheids enable astronomers to estimate vast distances

The details of a Cepheid's pulsation depend on the abundance of heavy elements in its atmosphere. The average luminosity of metal-rich Cepheids is roughly 4 times greater than the average luminosity of metal-poor Cepheids that have the same period. The difference in luminosity is a result of the higher mass of the metal-rich Cepheids compared to the metal-poor ones. Thus, there are two classes: **Type I Cepheids** (also called **δ Cephei stars**), which are the brighter, more massive, metal-rich stars, and **Type II Cepheids**, which are the dimmer, lower-mass, metal-poor stars. The period–luminosity relationships for both types of variables are shown in Figure 11-25.

Cepheids are important to astronomers because there is a direct relationship between a Cepheid's period of pulsation and its average luminosity. This correlation is called, appropriately enough, the **period–luminosity relationship**. We can determine the distance to a Cepheid in four steps. First, we observe its period. Second, we use the period–luminosity relationship, shown graphically in Figure 11-25, to determine its luminosity and, hence, its absolute magnitude (recall that the luminosity and absolute magnitudes are directly related to each other). Third, we observe its apparent magnitude.

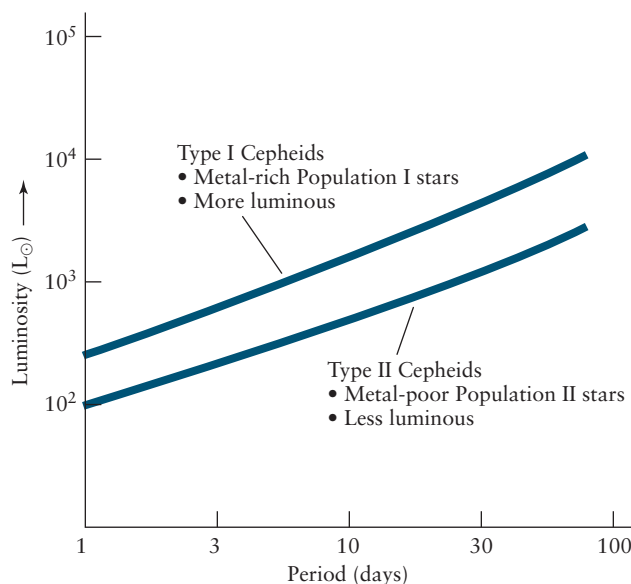


FIGURE 11-25 The Period–Luminosity Relationship The period of a Cepheid variable is directly related to its average luminosity: The more luminous the Cepheid, the longer its period and the slower its pulsations. Type I Cepheids (δ Cephei stars) are brighter, more massive, and more metal-rich stars than Type II Cepheids. The greater brightness of the Type I Cepheids is a result of their higher mass. (Data from H. C. Arp)

Finally, we use the distance–magnitude relationship in Discovery 10-2: The Distance–Magnitude Relationship to calculate its distance.

Intrinsically dim Type I Cepheid variables pulsate rapidly, with periods of 1 to 2 days, and have average brightness variations of a few hundred times our Sun’s luminosity. The most luminous of these Cepheids have the longest periods of all Cepheids, with variations occurring over 100 days and average brightness variations equaling 10,000 $L_{\odot}$. As they are so bright, we can see Cepheids far beyond the boundaries of our Milky Way

Galaxy. Because the changes in brightness of Cepheids can be seen at distances where other techniques for measuring distance, such as stellar parallax, fail, the period–luminosity relationship plays an important role in determining the overall size and structure of the universe. We will explore this application further in Chapter 14.

11-14 Globular clusters are bound groups of old stars

Many stars are found in groups, called **globular clusters**, so named because of their spherical shapes. A typical globular cluster, like the one shown in **Figure 11-26**, may contain up to a million stars in a volume 200 light-years across. Like open clusters, the stars in globular clusters all formed at about the same time. Unlike open clusters of young stars, however, globular clusters are gravitationally bound groups of old stars from which few stars escape. At least 157 globular clusters in our Galaxy are

Insight Into Science

Firm Foundations? To determine things far from our everyday size and timescales, scientists often must base their results on the predictions of complex theories, even before all of the testing and refinement of the theories are complete. An error anywhere along the chain of ideas leads to incorrect results. For example, using the Cepheid variable stars to measure distances requires combining several intermediate concepts, including believing that the period–luminosity relationship applies well to all Cepheids, believing that there are just two different types of Cepheids, and assuming that the peak luminosity depends only on the Cepheid’s distance from us (not correct if there are gas and dust between us and the star, as there often are). If any of these or myriad other assumptions are wrong, the calculated distances will be incorrect.

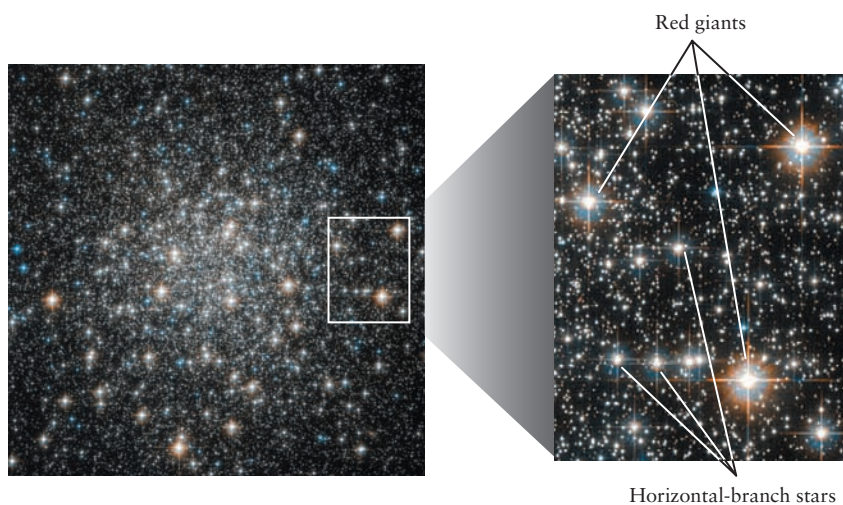


FIGURE 11-26 A Globular Cluster This cluster, M10, is about 85 light-years across and is located in the constellation Ophiuchus (the Serpent Holder), roughly 16,000 light-years from Earth. Most of the stars here are either red giants or blue horizontal-branch stars with both core helium fusion and hydrogen shell fusion. (©ESA/Hubble & NASA)

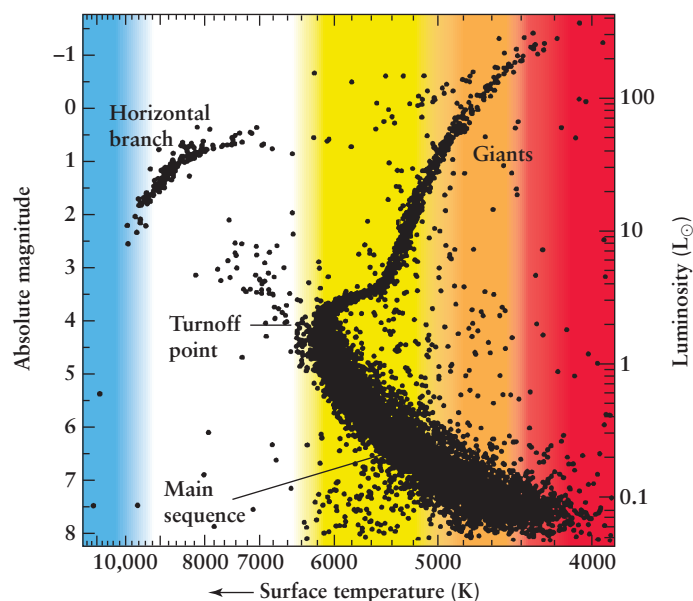


FIGURE 11-27 An H-R Diagram of a Globular Cluster

Each dot on this graph represents the absolute magnitude and surface temperature of a star in the globular cluster M55. Note that the upper half of the main sequence is missing. The horizontal-branch stars are stars of less than $2 M_{\odot}$ that recently experienced the helium flash in their cores and now exhibit core helium fusion and hydrogen shell fusion.

spherically distributed around the center of the Milky Way. Other galaxies also have them located in similar distributions.

Astronomers know that most globular clusters are old because they contain no high-mass main-sequence stars. (The young star cluster Westerlund 1 in our Galaxy and similar clusters elsewhere are apparently exceptions. These are discussed later in this section.) If you measure the luminosity and surface temperature of many stars in an older globular cluster and plot the data on a color-magnitude diagram (Figure 11-27), you will find that the upper half of the main sequence is missing. All of the high-mass main-sequence stars evolved long ago into giants, leaving behind only lower-mass, more slowly evolving main-sequence stars still undergoing core hydrogen fusion.

The H-R diagram of a globular cluster typically shows a roughly horizontal grouping of stars to the left of the center portion of the diagram (see Figure 11-27). These hot stars of less than $2 M_{\odot}$, called **horizontal branch stars** (see Figure 11-26), have undergone a helium flash and are now fusing helium in their cores. They have luminosities of about $50 L_{\odot}$. Eventually, these stars will move back toward the giant region as core helium fusion and hydrogen shell fusion devour their fuel.

As noted earlier, an H-R diagram of a cluster can also be used to determine the age of the cluster, assuming that all of the stars in the cluster formed at the same time. In the H-R diagram for a young open cluster (review Figure 11-17b), almost all of the hot massive stars are still on the main sequence. As a cluster ages, like the Pleiades, stars begin to leave the main sequence (see Figure 11-2b). The high-mass, high-luminosity stars

are the first to become giants. We can determine the ages of globular clusters in the same way, since it appears that all the stars in a given cluster form at the same time.

Over time, the main sequence in a cluster gets shorter and shorter. The top of the surviving portion of the main sequence is called the **turnoff point** (see Figure 11-27). Stars at the turnoff point are just beginning to exhaust the hydrogen in their cores. The theory of stellar evolution enables us to predict how long stars with different masses last on the main sequence before moving toward the giant stage. Therefore, the age of the stars currently at the turnoff point is the age of the cluster (recall Table 11-2). Consider the plot for the globular cluster M55, shown in Figure 11-27. In M55, $0.9 M_{\odot}$ stars are just leaving the main sequence. Therefore, according to Table 11-2, the cluster's age is approximately 12.5 billion years.

Using data collected from the *Hipparcos* satellite, Figure 11-28a is a composite diagram of isolated and cluster stars, reflecting the overall structure of the H-R diagram. Using data from stellar evolution theory as presented in Table 11-2, the ages of star clusters can be estimated from the turnoff points, such as those depicted in Figure 11-28b. The evolution of a cluster is summarized in Figure 11-28c–j.

Until recently, astronomers believed that all globular clusters were very old, with turnoff points as just described. However, analysis of some clusters, especially Westerlund 1, discovered in 2004, suggests that globular clusters are still forming. Westerlund 1, located in our Galaxy, contains roughly half a million stars in a volume

Focus Question 11-13

What are typical periods of Cepheid variable stars?

Focus Question 11-14

Which are older, stars in open clusters or stars in globular clusters?

only 6 light-years across, including at least 200 O and B stars. The presence of such high-mass stars implies that Westerlund 1 is only about 5 million years old. This cluster is called a *superstar cluster*. Such clusters are believed to be early stages in the lives of globular clusters. Superstar clusters have been observed in other galaxies (see, for example, the figure opening this chapter).

While most globular clusters contain many of the oldest stars in our Galaxy, the youngest stars in the Milky Way are found in its open clusters. Unlike globular clusters, which are spherically distributed around the center of the Milky Way, our Galaxy's open clusters lie in the plane of the Galaxy, that is, along the band of light sweeping majestically across the night sky.

Stars in open clusters are said to be metal-rich, because their spectra contain many prominent spectral lines of heavy elements. This material originally came from stars that exploded long ago, enriching the interstellar gases with the heavy elements formed in their cores. The young, open clusters are therefore formed from the debris of older generations of stars. The Sun is an example of a young, metal-rich star that was

likely formed in an open cluster. Such stars are called **Population I stars**.

Older globular clusters are generally located above or below the plane of our Galaxy. Because their spectra show only weak lines of heavy elements, the ancient stars in these clusters are said to be metal-poor. They were created long ago from gases that had not yet been substantially enriched with heavy elements and are called **Population II stars**. **Figure 11-29** compares the spectra of a Population II star and the Sun. As we saw for Cepheid variable stars (see Figure 11-25), differences in chemistries can affect stellar evolution.

11-15 Mass transfer in close binary systems can produce unusual double stars

Some stars substantially change their masses over time and thereby alter the patterns of their evolution. This occurs for stars in binary systems that are sufficiently close to each other (see Chapter 10). In the mid-1800s, French mathematician Édouard Roche pointed out that the atmospheres of two stars in a binary system

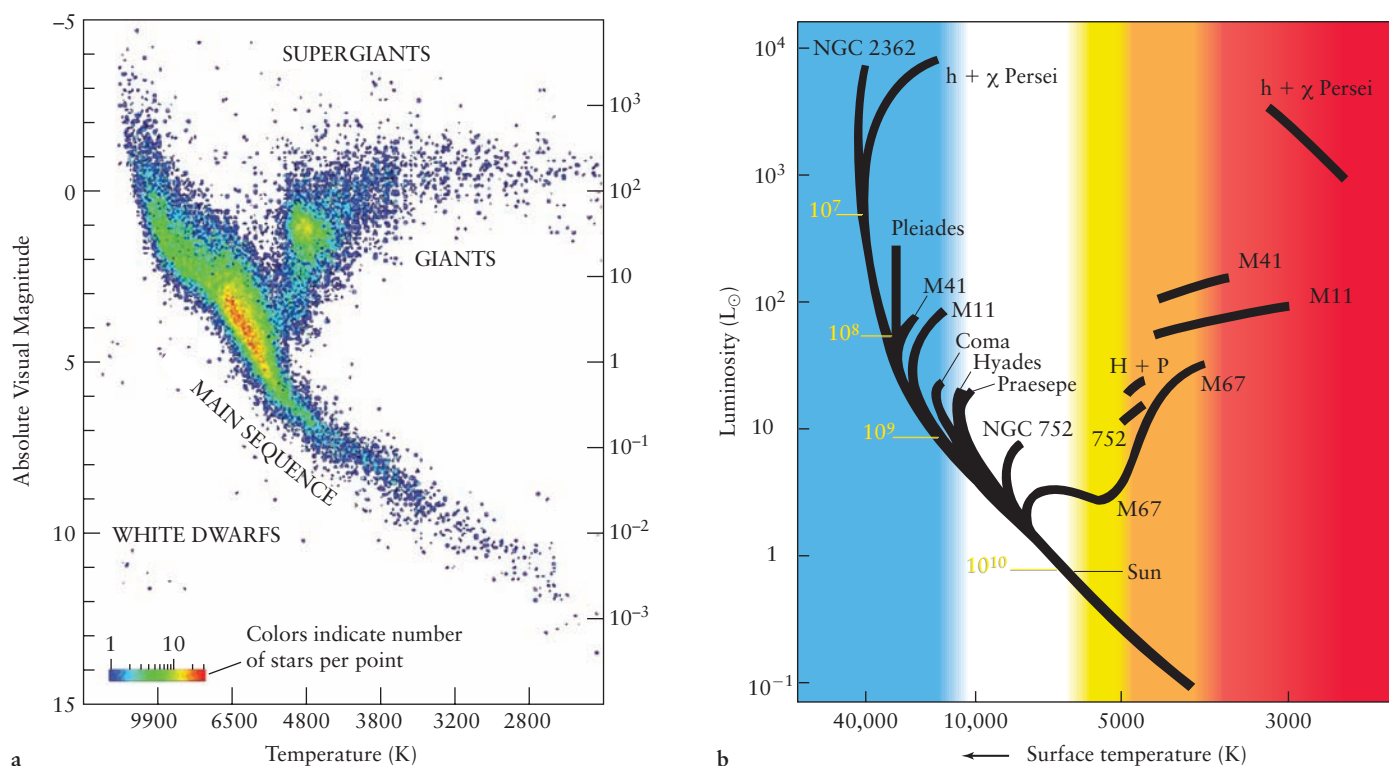


FIGURE 11-28 Structure of the H-R Diagram (a) Data taken by the *Hipparcos* satellite placed 41,453 stars more precisely on the H-R diagram than any previous observations. This figure shows the overall structure of the H-R diagram. The thickness of the main sequence is due in large part to stars of different ages turning off the main sequence at different places, as

shown in (b). (b) The black bands indicate where data from various star clusters fall on the H-R diagram. The ages of turnoff points (in years) are listed in yellow alongside the main sequence. The age of a cluster can be estimated from the location of the turnoff point, where the cluster's most massive stars are just now leaving the main sequence.

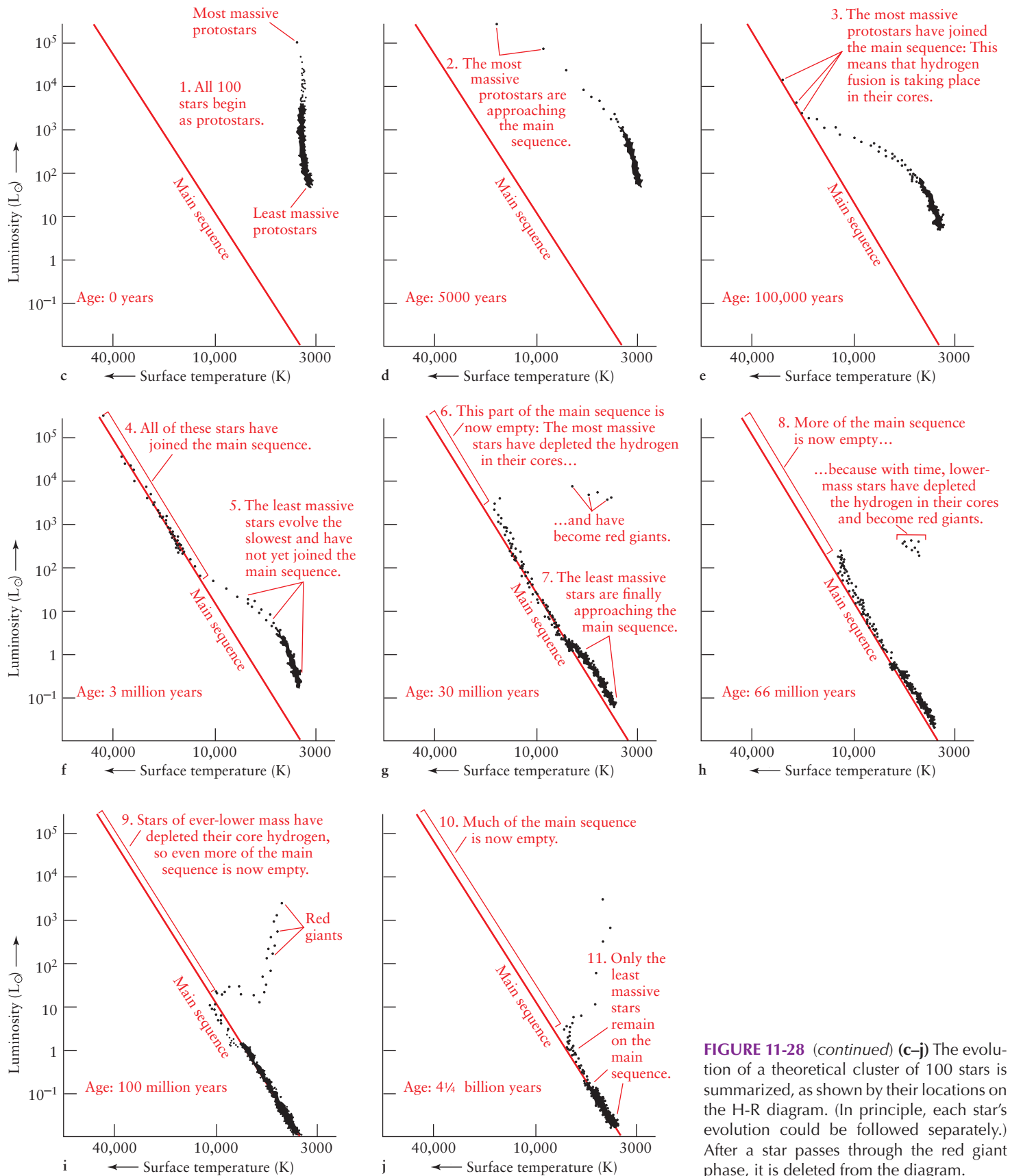


FIGURE 11-28 (continued) (c–j) The evolution of a theoretical cluster of 100 stars is summarized, as shown by their locations on the H-R diagram. (In principle, each star's evolution could be followed separately.) After a star passes through the red giant phase, it is deleted from the diagram.

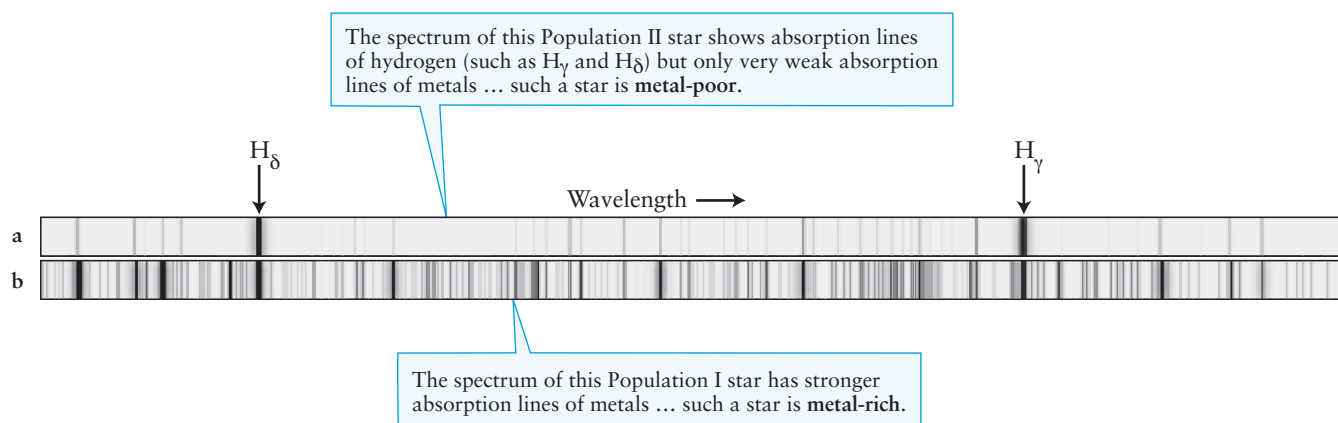


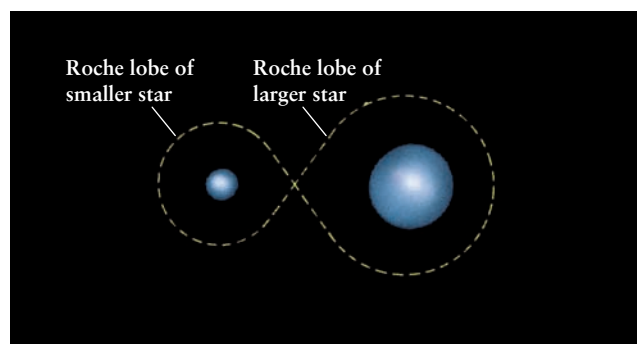
FIGURE 11-29 Spectra of a Metal-Poor and a Metal-Rich Star These spectra compare (a) a metal-poor (Population II) and (b) a metal-rich (Population I) star (the Sun) of the same surface temperature. Numerous spectral lines prominent in the solar spectrum are caused by elements heavier than hydrogen

and helium. Note that corresponding lines in the metal-poor star's spectrum are weak or absent. Both spectra cover a wavelength range that includes two strong hydrogen absorption lines, labeled H_γ (410 nm) and H_δ (434 nm).

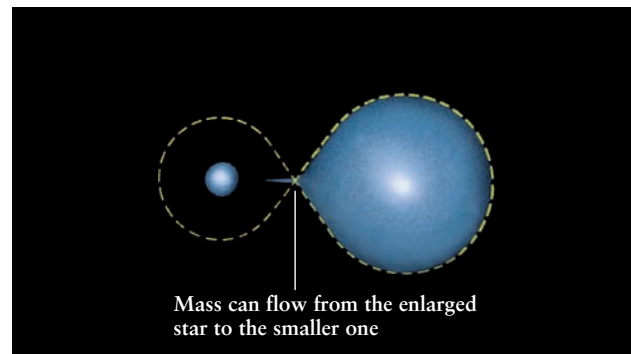
must remain within a pair of teardrop-shaped regions that surround the stars. Otherwise, the gas escapes from the star of its origin, either transferring to the other star, where the teardrops make contact, or escaping completely from the binary in the opposite direction. In cross section, these so-called **Roche lobes** take on a figure-eight

shape (dashed lines in **Figure 11-30**). The more massive star is always located inside the larger Roche lobe.

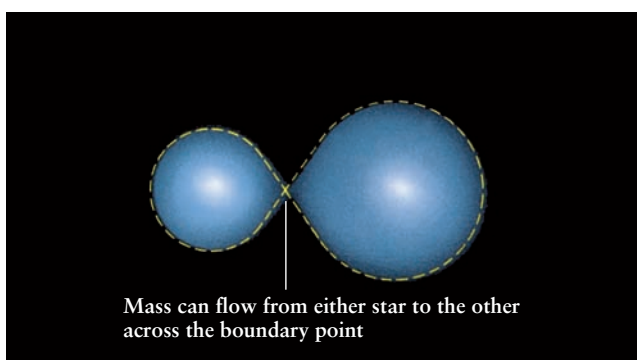
The stars in many binaries are so far apart that, even during their giant stages, the stars' surfaces remain well inside their Roche lobes. Other than orbiting one another, each star in such systems lives out its life cycle as



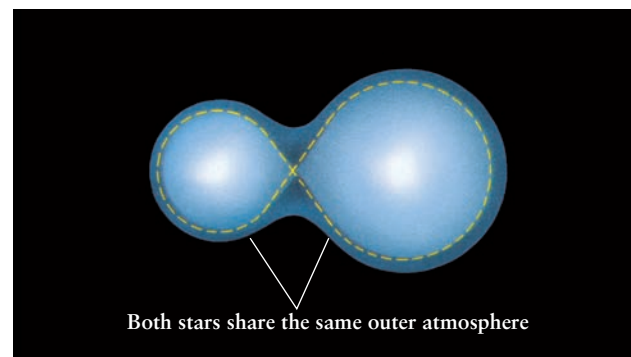
a Detached binary: Neither star fills its Roche lobe.



b Semidetached binary: One star fills its Roche lobe.



c Contact binary: Both stars fill their Roche lobes.



d Overcontact binary: Both stars overflow their Roche lobes.

FIGURE 11-30 Detached, Semidetached, Contact, and Overcontact Binaries (a) In a detached binary, neither star fills its Roche lobe. (b) If one star fills its Roche lobe, the binary is semidetached. Mass transfer is often observed in semidetached

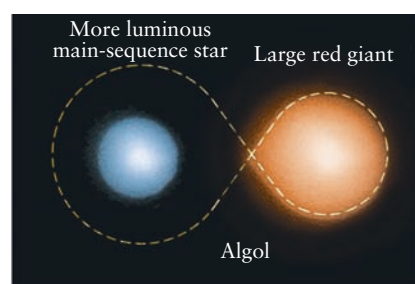
binaries. (c) In a contact binary, both stars fill their Roche lobes. (d) The two stars in an overcontact binary both overflow their Roche lobes. The two stars actually share the same outer atmosphere.

if it were single and isolated, and the system is referred to as a **detached binary** (see Figure 11-30a). If two stars are relatively close together, however, one star may fill or overflow its Roche lobe as it expands into the giant phase. This system is a **semidetached binary**, and gases then flow across the point where the two Roche lobes touch and fall onto the companion star (Figure 11-30b). When both stars completely fill their Roche lobes, the system is a **contact binary**, because the two stars actually touch and exchange gas (Figure 11-30c). It is quite unlikely, however, that both stars will exactly fill their Roche lobes at the same time. It is more likely that they overflow their lobes, giving rise to a common atmospheric envelope. Such a system is an **overcontact binary** (Figure 11-30d).

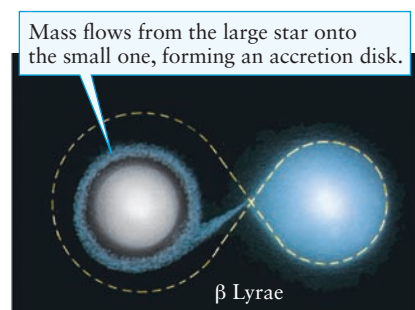
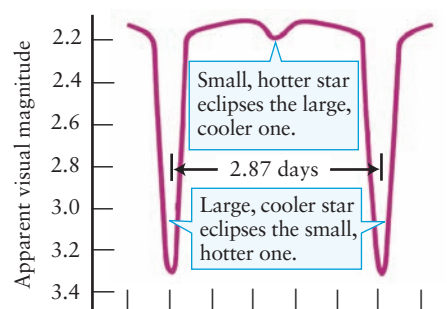
Semidetached and contact binaries are easiest to detect if they are also eclipsing binaries (recall Figure 10-13). Their light curves have a distinctly rounded appearance, caused by these tidally distorted egg-shaped stars. The

eclipsing binary β (beta) Persei, commonly called Algol (from the Arabic term for “demon”), is a semidetached binary that can easily be seen with the naked eye in the constellation Perseus. From Algol’s light curve (Figure 11-31a), astronomers have determined that the binary contains a star that fills its Roche lobe. Sometime in the past, as it expanded and became a giant, this star dumped a significant amount of gas onto its companion.

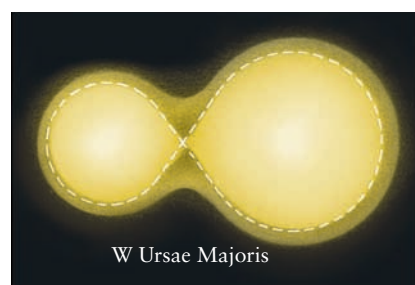
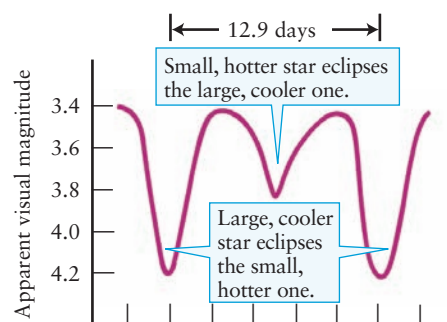
Mass transfer is still occurring in a semidetached eclipsing binary, called β Lyrae, in the constellation Lyra. Like β Persei, β Lyrae contains a giant that fills its Roche lobe (Figure 11-31b). For many years, astronomers were puzzled as to why the detached companion star in β Lyrae is severely underluminous, contributing virtually no light at all to the visible radiation coming from the system. Furthermore, the spectrum of β Lyrae contains unusual features, some of which are consistent with gas flowing between the stars and around the system as a whole.



a A semidetached binary



b A semidetached binary with mass transfer



c An overcontact binary

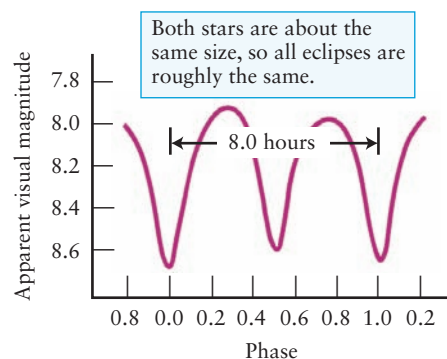


FIGURE 11-31 Three Close Binaries

Sketches of light curves for three eclipsing binaries are shown. The phase denotes the fraction of the orbital period from one primary minimum to the next. (a) Algol, also known as β Persei, is a semidetached binary. The deep eclipse occurs when the giant star (right) blocks the light from the smaller but more luminous main-sequence star. (b) β Lyrae is a semidetached binary in which mass transfer has produced an accretion disk that surrounds the detached star. This disk is so thick and opaque that it renders the secondary star almost invisible. (c) W Ursae Majoris is an overcontact binary. Both stars therefore share their outer atmospheres. The short, 8-hour period of this binary indicates that the stars are very close to each other.

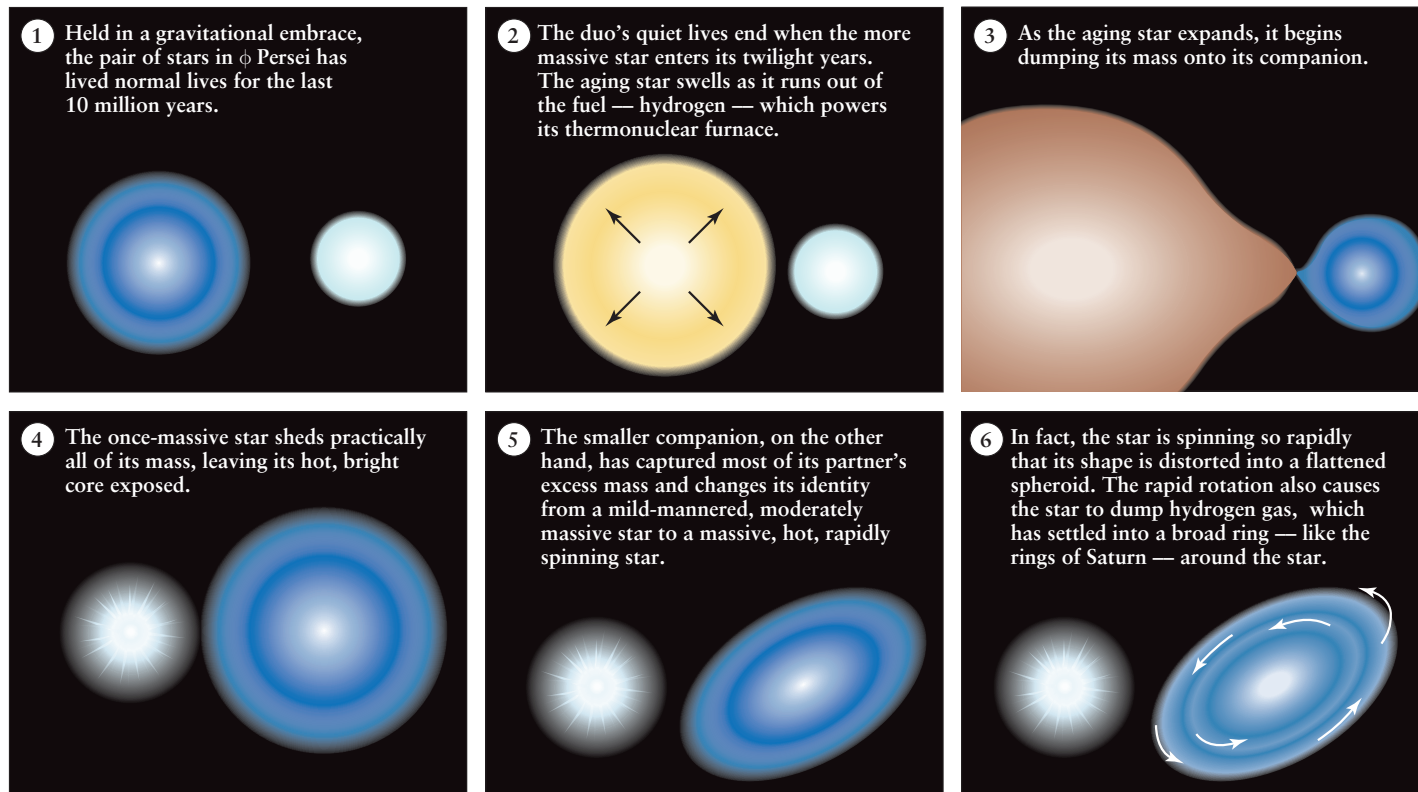


FIGURE 11-32 Mass Exchange Between Close Binary Stars This sequence of drawings shows how close binary stars can initially be isolated but, as they age, grow and exchange

mass. Such mass exchange leads to different fates than if the same stars had evolved in isolation.

The β Lyrae system was explained in 1963, when the Chinese American astronomer Su-Shu Huang proposed that the underluminous star in β Lyrae is enveloped in a huge **accretion disk** of gas captured from its bloated companion. The disk is so large and thick that it completely enshrouds the secondary star, making it impossible to observe at visible wavelengths. Observations indicate that the primary star is overflowing its Roche lobe, with gases streaming onto the disk at the rate of $1 M_{\odot}$ per hundred thousand years.

The fate of a semidetached system like β Persei or β Lyrae depends primarily on how fast its stars evolve. If the detached star expands to fill its Roche lobe while the companion star fills its own Roche lobe, then the result is an overcontact binary. An example is W Ursae Majoris, in which two stars share the same photosphere (Figure 11-31c).

Theories of the evolution of binary systems that transfer mass reveal that some remarkable transformations are possible, such as the case of ϕ (phi) Persei. As shown in **Figure 11-32**, the more massive star in a binary

can actually be transformed into the lower-mass star, and vice versa! When stars in such systems gain mass, their increased gravitational force increases their rate of fusion, which increases the rate at which they evolve. Conversely,

the stars that lose mass begin evolving more slowly than before their mass loss. Observations in 2012 revealed that 70% of O-type stars are in close binary systems. In Chapter 12 we will see that mass transfer produces some of the most extraordinary objects in the sky.

SUMMARY OF KEY IDEAS

Protostars and Pre–Main-Sequence Stars

- Enormous, cold clouds of gas and dust, called giant molecular clouds, are scattered about the disk of the Galaxy.
- Star formation begins when gravitational attraction causes clumps of gas and dust, called protostars, to coalesce in Bok globules within a giant molecular cloud. As a protostar contracts, its matter begins to heat and glow. When the contraction slows down, the protostar becomes a pre–main-sequence star. When the pre–main-sequence star's core temperature becomes high enough to begin hydrogen fusion and stop contracting, it becomes a main-sequence star.
- The most massive pre–main-sequence stars take the shortest time to become main-sequence stars (O and B stars), while the least massive pre–main-sequence stars take the longest time.
- In the final stages of pre–main-sequence contraction, when hydrogen fusion is about to begin in the core, the

Focus Question 11-15

In which type of binary system are the two stars most actively exchanging mass?

pre-main-sequence star may undergo vigorous chromospheric activity that ejects large amounts of matter into space. G, K, and M stars at this stage are called T Tauri stars.

- A collection of a few hundred or a few thousand newborn stars formed in the plane of the Galaxy is called an open cluster. Stars escape from open clusters, most of which eventually dissipate.

Main-Sequence and Giant Stars

- The Sun has been a main-sequence star for 4.6 billion years and will remain so for about another 5 billion years. Less massive stars than the Sun evolve more slowly and have longer main-sequence lifetimes. More massive stars than the Sun evolve more rapidly and have shorter main-sequence lifetimes.
- Low-mass (between $0.08 M_{\odot}$ and $0.4 M_{\odot}$) main-sequence stars convert all of their mass into helium and then stop fusing. Their lifetimes last hundreds of billions of years, so none of these stars has yet left the main sequence.
- Core hydrogen fusion ceases when hydrogen in the core of a main-sequence star with $M > 0.4 M_{\odot}$ is gone, leaving a core of nearly pure helium surrounded by a shell where hydrogen fusion continues. Hydrogen shell fusion adds more helium to the star's core, which contracts and becomes hotter. The outer atmosphere expands considerably, and the star becomes a giant.
- When the central temperature of a giant reaches about 100 million K, the thermonuclear process of helium fusion begins. This process converts helium to carbon and then to oxygen. In a massive giant, helium fusion begins gradually. In a less massive giant, it begins suddenly in a process called the helium flash.
- Giants undergo extensive mass loss, sometimes producing shells of ejected material that surround the entire star.
- Relatively young stars are metal-rich (Population I); ancient stars are metal-poor (Population II).

Clusters of Stars

- Groups of between a few hundred and a few thousand stars, formed together from a single interstellar cloud in the disk of our Galaxy, are called open clusters.
- Stars in open clusters go their separate ways.
- Groups of hundreds of thousands to millions of stars formed together from a common interstellar cloud are called globular clusters.
- Stars in globular clusters remain bound together.
- The age of an open or globular star cluster can be estimated by plotting its stars on an H-R diagram. The upper portion of the main sequence disappears first, because more massive main-sequence stars become giants before low-mass stars do. Therefore, the age of the stars now leaving the main sequence is the cluster's age.

Variable Stars

- When a star's evolutionary track carries it through a region called the instability strip in the H-R diagram, the star becomes unstable and begins to pulsate.
- RR Lyrae variables are low-mass, pulsating variables with short periods. Cepheid variables are higher-mass, pulsating variables exhibiting a regular relationship between the period of pulsation and luminosity.
- Mass can be transferred from one star to another in close binary systems. When this occurs, the evolutionary paths of the two stars change.

WHAT DID YOU THINK?

- 1 *How do stars form?* Each star forms from the collective gravitational attraction of a clump of gas and dust usually inside a giant molecular cloud.
- 2 *Are stars still forming today? If so, where?* Yes. Astronomers have seen stars that have just arrived on the main sequence, as well as infrared images of gas and dust clouds in the process of forming stars. Most stars in the Milky Way form in giant molecular clouds in the disk of the Galaxy.
- 3 *Do more massive stars shine longer than less massive ones? Explain your reasoning.* No. Lower-mass stars last longer because the lower gravitational force inside them causes fusion to take place at much slower rates compared to the fusion inside higher-mass stars. These latter stars therefore use up their fuel more rapidly than do lower-mass stars.
- 4 *When stars like the Sun stop fusing hydrogen into helium in their cores, do the stars get smaller or larger?* They get larger. Such stars start fusing hydrogen into helium outside their cores. This new fusion, closer to the star's surface, is able to push the star's outer layers out farther than they had been before.

Review Questions

1. Consider a star behind a cloud of interstellar gas and dust and at rest as seen from our perspective. Which of the following statements is correct?
 - a. The star appears brighter than it would if the cloud were not present.
 - b. The star appears to be blueshifted.
 - c. The star appears redder than it would if the cloud were not present.
 - d. The star would be invisible at all wavelengths.
 - e. The star would always appear green.
2. What is the lowest mass that a star can have on the main sequence?

a. There is no lower limit.	d. $0.4 M_{\odot}$
b. $0.003 M_{\odot}$	e. $2.0 M_{\odot}$
c. $0.08 M_{\odot}$	

3. What is the source of energy that enables a main-sequence star to shine?
 - a. Friction between its atoms
 - b. Fusion of hydrogen in a shell that surrounds the core
 - c. Fusion of helium in its core
 - d. Fusion of hydrogen in its core
 - e. Burning of gases on its surface
4. What are giant molecular clouds, and what role do they play in star formation?
5. Why are low temperatures necessary for dense cores to form and contract into protostars?
6. Why do thermonuclear reactions not occur on the surface of a main-sequence star?
7. What is an evolutionary track, and how do such tracks help us interpret the H-R diagram?
8. Draw the pre-main-sequence evolutionary track of the Sun on an H-R diagram. Briefly describe what occurred throughout the solar system at various stages along this track. (You may find it useful to review Chapter 10.)
9. On what grounds are astronomers able to say that the Sun has about 5 billion years remaining in its main-sequence stage?
10. What will happen inside the Sun 5 billion years from now when it begins to evolve into a giant?
11. How is the evolution of a main-sequence star with less than $0.4 M_{\odot}$ fundamentally different from that of a main-sequence star with more than $0.4 M_{\odot}$?
12. Draw the post-main-sequence evolutionary track of the Sun on an H-R diagram up to the point when the Sun becomes a helium-fusing giant. Briefly describe what might occur throughout the solar system as the Sun undergoes this transition.
13. What does it mean when an astronomer says that a star “moves” from one place to another on an H-R diagram?
14. What is the helium flash and what causes it?
15. Explain how and why the turnoff point on the H-R diagram of a cluster is related to the cluster’s age.
16. Why do astronomers believe that most globular clusters are made of old stars?
17. What are Cepheid variables, and how are they related to the instability strip?
18. What occurs in Cepheid stars that is analogous to the vapor raising the lid on a pot of boiling water?

19. What are RR Lyrae variables, and how are they related to the instability strip?
20. What is a Roche lobe, and what is its significance in close binary systems?
21. What are the differences between detached, semi-detached, contact, and overcontact binaries?

Advanced Questions

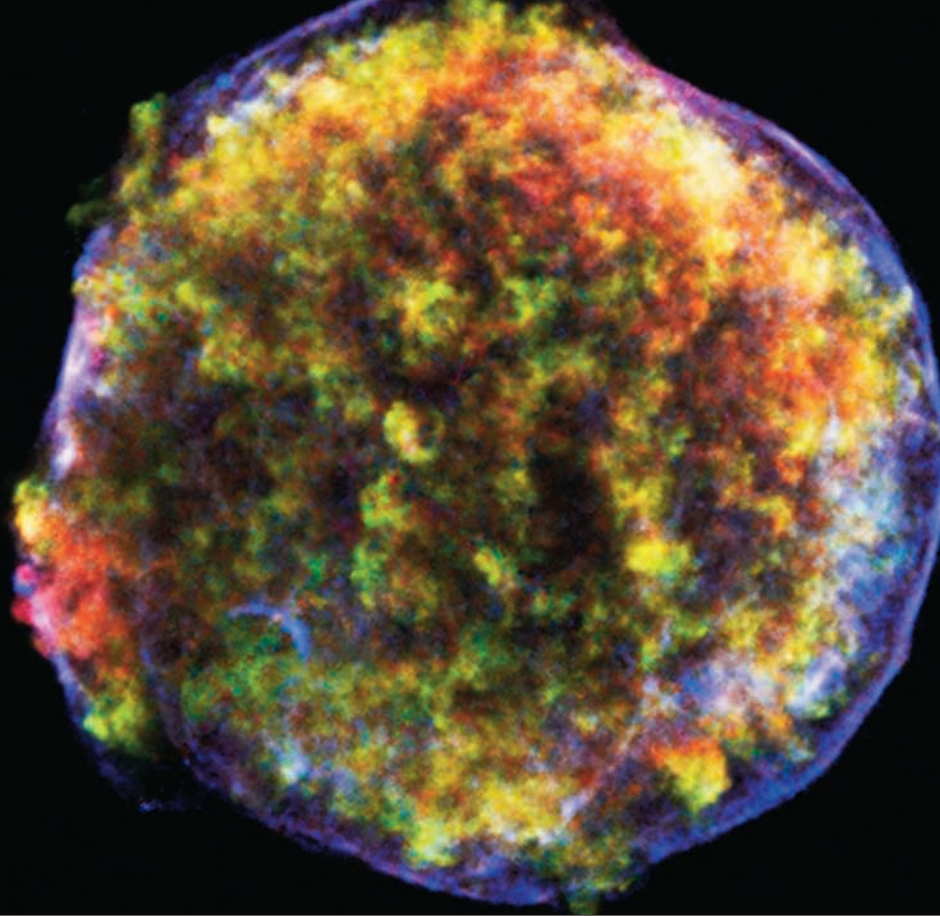
22. How is a degenerate gas different from an ordinary gas?
23. If you took a spectrum of a reflection nebula, would you see absorption lines, emission lines, or no lines? Explain your answer.
24. Why is it useful to plot the *apparent* magnitudes of stars in a single cluster on an H-R diagram?
25. What might happen to the massive outer planets when the Sun becomes a giant?
26. Why does a shock wave from a supernova produce relatively few high-mass O and B stars compared to the number of low-mass A, F, G, K, and M stars produced?
27. How would you distinguish a newly formed protostar from a giant, given that they occupy the same location on the H-R diagram?
28. What observational consequences would we find in H-R diagrams for star clusters as a result of the universe having a finite age? Could we use these consequences to establish constraints on the possible age of the universe? Explain your answers.

Discussion Questions

29. Discuss the possibility of life-forms and biological processes occurring in giant molecular clouds. In what ways might conditions favor or hinder biological evolution?
30. Is there any evidence that Earth has ever passed through a star-forming region in space?

Got It?

31. Which “last” longer, high-mass or low-mass stars? Explain your reasoning.
32. Is Jupiter a “failed” star? Why or why not?
33. Which of the two stars, if either, is at the center of a binary star system? Explain your reasoning.
34. Did the Sun form shortly after the universe came into existence? Explain your answer.
35. Are stars still forming today? Explain your answer.



CHAPTER 12

RIVU X G This X-ray image of Tycho's supernova, first seen as a visible-light object by Tycho Brahe in 1572, was taken in 2003 by the Chandra X-ray Observatory. Gas and dust with temperatures in the millions of kelvins (shown in red and green) are expanding outward at about 10 million km/hr (6.2 million mi/hr), following a shell of high-energy electrons (blue). (NASA/CXC/Rutgers/J. Warren and J. Hughes et al.)

The Deaths and Remnants of Stars

WHAT DO YOU THINK?

- 1 Will the Sun someday cease to shine brightly? If so, how will this occur?
- 2 What is a nova? How does it differ from a supernova?
- 3 What are the origins of the carbon, silicon, oxygen, iron, uranium, and other heavy elements on Earth?
- 4 What are cosmic rays? Where do they come from?
- 5 What is a pulsar?
- 6 Are black holes empty holes in space? If not, what are they?
- 7 Does a black hole have a solid surface? If not, what is at its surface?
- 8 What power or force enables black holes to draw things into themselves?
- 9 How close to a black hole do you have to be for its special effects to be apparent?
- 10 Can you use black holes to travel to different places in the universe?
- 11 Do black holes last forever? If not, what happens to them?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

In this chapter we learn how the later stages of stellar evolution are significantly different for stars with different masses. Some of them stop evolving by relatively mild emissions of their outer layers, while others have spectacular finales. Then we will explore the properties of the remnants left over from the stars.

In this chapter you will discover

- what happens to stars when core helium fusion ceases
- how heavy elements are created
- the characteristics of the end of stellar evolution
- why some stars go out relatively gently and others go out with a bang
- the incredible density of the matter in neutron stars and how these objects are observed
- that Einstein's theory of general relativity predicts the existence of regions of space and time that are severely distorted by the extremely dense matter they contain
- that space and time are not separate entities
- how black holes arise
- the surprisingly simple theoretical properties of black holes
- that X-rays and jets of gas are created near many black holes
- the fate of black holes
- the unsurpassed energy emitted by gamma-ray bursts

INTERMEDIATE-MASS ($0.4 M_{\odot}$ – $8 M_{\odot}$) STARS AND PLANETARY NEBULAE

We have seen that red dwarf stars, with masses less than $0.4 M_{\odot}$, will never get to the giant phase of stellar evolution. Rather, they will eventually convert all their mass to helium, stop fusing, and then just cool off. We will

now explore the fates of stars greater than $0.4 M_{\odot}$ as they proceed along through the giant phase and beyond.

In Chapter 11, we discovered that when hydrogen fusion in the shell surrounding the core first begins, the energy from it causes a star to expand and become a giant. During this process, intermediate-mass stars move over to, and ascend, the giant branch on the H-R diagram for the first time (Figure 12-1a; see also Figure 11-23b).

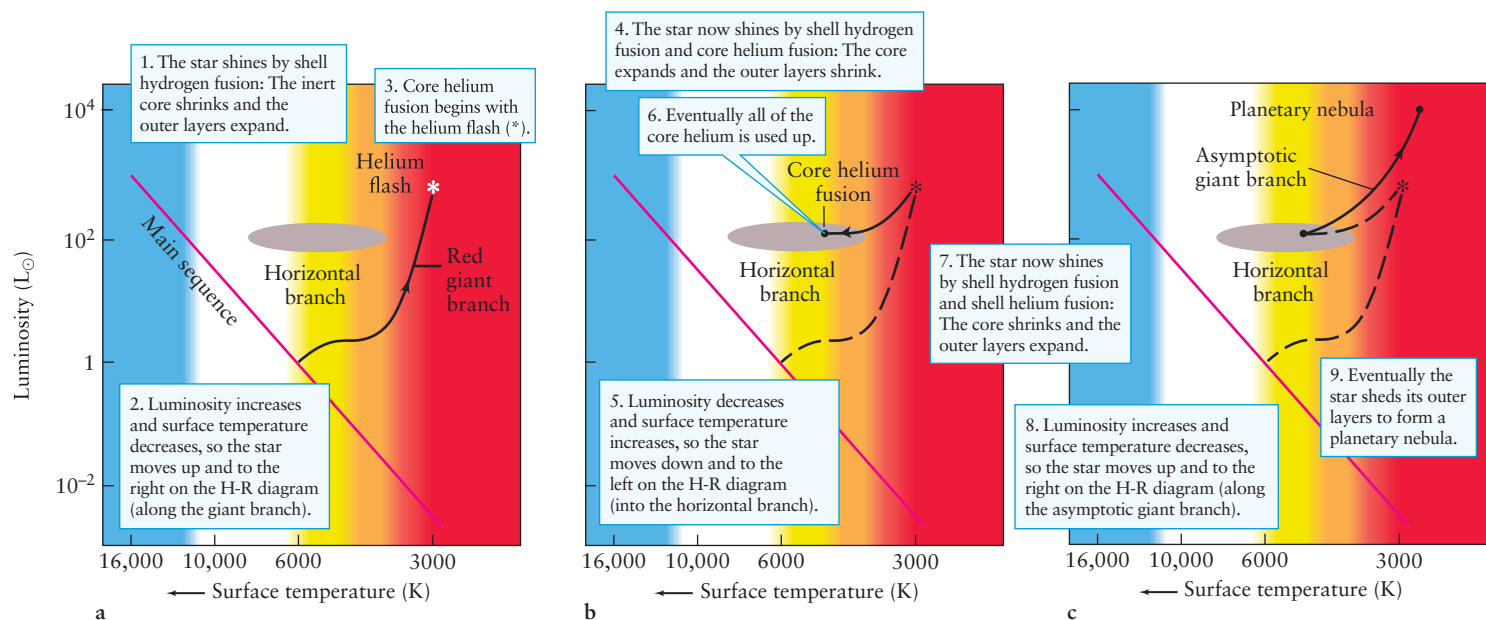


FIGURE 12-1 Post-Main-Sequence Evolution of Intermediate-Mass Stars (a) A typical evolutionary track on the H-R diagram as a star makes the transition from the main sequence to the giant phase. The asterisk (*) shows the helium flash occurring in an intermediate-mass star. (b) After the helium flash, the star converts its helium core into carbon and oxygen. While doing so, its core expands again, de-

creasing shell fusion. As a result, the star's outer layers recontract. (c) After the helium core is completely transformed into carbon and oxygen, the core recollapses, and the outer layers expand again, powered up the asymptotic giant branch by hydrogen shell fusion and helium shell fusion.

As this happens, mass is expelled into space in the form of stellar winds, which reduce the masses of these stars. Then core helium fusion begins, with stars of less than about $2 M_{\odot}$ undergoing a core helium flash in the giant stage (Figure 12-1a). Recall also that stars with more than $2 M_{\odot}$ begin helium fusion more gradually. In either case, after core fusion begins again, stars shrink and move onto the *horizontal branch* (Figure 12-1b).

The destinies of stars depend on their masses, and we have two mass ranges to consider: $0.4 M_{\odot}$ to $8 M_{\odot}$ (*intermediate-mass stars*) and more than $8 M_{\odot}$ (*high-mass stars*).

12-1 Intermediate-mass stars become supergiants before expanding into planetary nebulae

The complete transformation of the helium core of an intermediate-mass star into carbon occurs on the horizontal branch, after which core fusion again temporarily ceases. Calculations reveal that carbon (and, later, oxygen) atoms require a temperature of at least 600 million K to fuse. Because the core of an intermediate-mass giant on the horizontal branch only reaches about 200 million K, fusion of these elements in the core does not occur. Photon production in the core therefore decreases when its helium fuel is nearly used up and, as a result, the inner regions of the star again contract. Just outside the core, the hydrogen shell fusion has created a shell of helium. As the core contracts, this helium follows it inward and is thereby compressed and heated until it reaches 100 million K. At that temperature, he-

lium shell fusion begins in what used to be the realm of hydrogen shell fusion. This shell is itself surrounded by a new hydrogen-fusing shell (Figure 12-2). All of this takes place within a volume roughly the size of Earth.

Once helium shell fusion commences, the new outpouring of energy again pushes out the outer envelope of the star. An intermediate-mass star thus leaves the horizontal branch and ascends the H-R diagram for a second, and final, time (Figure 12-1c). Powered by the fusion from two shells, such an **asymptotic giant branch star**, or **AGB star**, becomes brighter than ever before. When the Sun moves along the asymptotic giant branch, it will expand to a radius of about 1 AU, enveloping Earth in its tenuous outer layers (Figure 12-2). The fate of the Earth as the Sun evolves in the future is an area of intense discussion among scientists.

In comparison, an $8 M_{\odot}$ star will have a diameter as big as the orbit of Mars and shine with the luminosity of $10^4 L_{\odot}$. At the top of the asymptotic giant branch, all these stars are so bright that they are classified as low-temperature, red supergiants (see Figure 12-1c).

AGB stars are destined to self-destruct. Giant stars of all spectral types emit strong stellar winds. Lower-mass stars expel their outer layers more slowly than higher-mass stars, but both types of stars reduce their masses significantly from what they were on the main sequence. During their initial ascent up the giant branch, stars can lose as much as 30% of their masses. On the asymptotic giant branch, stars lose even more mass, often surrounding themselves with thickening cocoons of gas and dust. A star of the Sun's mass loses about $10^{-5} M_{\odot}$ per year at this stage, eventually dumping half of its mass back into space.

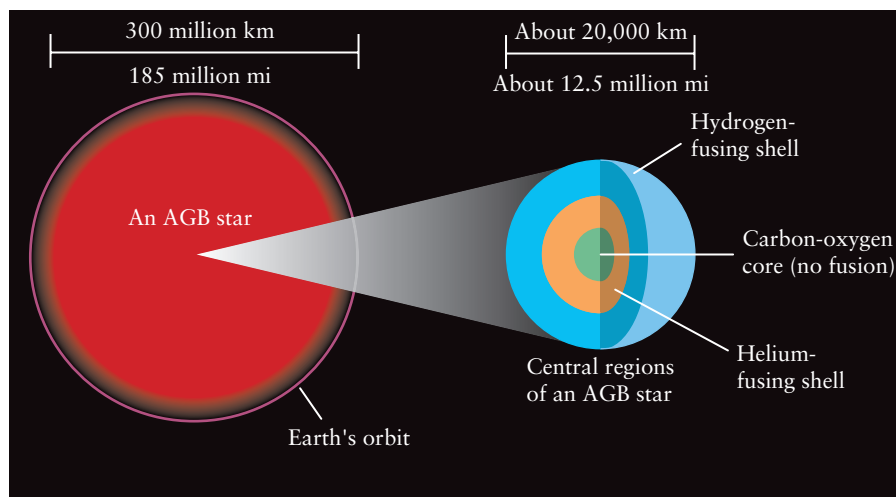


FIGURE 12-2 The Structure of an Old Intermediate-Mass Star Near the end of its life, an intermediate-mass star, like the Sun, travels up the AGB and becomes a red supergiant. (The Sun will eventually be about as large as the diameter of Earth's

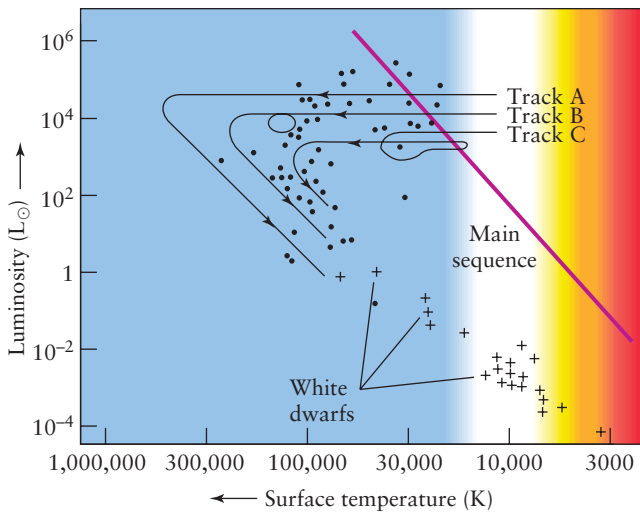
orbit.) The star's inert core, its helium-fusing shell, and its hydrogen-fusing shell are contained within a volume roughly the size of Earth. The inner layers are not shown to scale here.

This mass loss limits the force of gravity available to compress the star’s core and the regions of shell fusion. A fine balance is struck: The intermediate-mass AGB star is compressed until its core of carbon and oxygen becomes degenerate, meaning that the electrons provide a growing repulsive force that stops its contraction (as discussed in Section 11-10). However, even in the most massive of these intermediate-mass stars (now much less than their original $8\text{ }M_{\odot}$ due to their stellar winds), the core temperature cannot reach the 600 million K necessary to fuse carbon or oxygen. Therefore, no further core fusion occurs.

The final stage through which an intermediate-mass star passes begins with a “thermal runaway” (meaning a rapid rise in temperature) in the helium shell, just like the helium flash in its core earlier in its life (described in Section 11-10). This process occurs because the triple alpha process (see Section 11-11) is extremely sensitive to temperature and, as the temperature goes up slightly, the fusion rate skyrockets. The increase in energy output from the thin helium-fusing shell, called a **helium shell flash**, expands the star, thereby briefly decreasing the temperature in the shell and slowing the rate of fusion.

A star in this mass range undergoes several helium shell flashes as its helium shell thickens. During each flash, the helium shell’s energy output jumps a thousand-fold. These brief outbursts are separated by relatively quiet intervals lasting about 100,000 years, during which the helium shell gradually becomes thicker. The energy from the increased fusion during the helium shell flashes causes the star’s outer layers to expand and, therefore, to cool further. Eventually, the outer gases are sufficiently cool so that electrons and ions in them can recombine. This process is exactly the opposite of ionization (discussed in Section 3-16). Whereas ionization requires an electron to absorb a photon, recombination forces an electron to emit a photon.

Just as the collisions of particles with each other create pressure (force acting over an area), the collisions of photons with matter also create pressure. The impacts of photons emitted during recombination, along with the photons from the helium flashes and the photons normally flowing outward from fusion, generate enough pressure to eject more and more of the star’s outer layers into space. Eventually, the mass of the star decreases so much that there is not enough pressure to sustain any shell fusion. As the ejected material expands and cools, some of it condenses to form dust grains. When enough of the gas has left the star for the core to be visible, the expanding dust and gases are considered to be a **planetary nebula** (see Figure 12-1c). As the outer layers of the star are shed, an increasingly hot interior is revealed, and the star moves to the left



Evolutionary track	Mass ($M_{\odot}$)		
	Red supergiant	Ejected nebula	White dwarf
A	3.0	1.8	1.2
B	1.5	0.7	0.8
C	0.8	0.2	0.6

FIGURE 12-3 Evolution from Red Supergiants to White Dwarfs The evolutionary tracks of three intermediate-mass supergiants are shown as they eject planetary nebulae. The table gives their masses as supergiants, the amount of mass they lose as planetary nebulae, and their remaining (white dwarf) masses. The loops indicate periods of instability and adjustment for the white dwarfs. The dots on this graph represent the central stars of planetary nebulae whose surface temperatures and luminosities have been determined. The crosses are white dwarfs for which similar data exist.

across the H-R diagram (Figure 12-3). Including the mass lost prior to the planetary nebula phase, intermediate-mass stars lose as much as 80% of their mass by shedding their outer layers.

Insight Into Science

Names Often Don’t Change as Understanding Improves Planetary nebulae have *nothing* to do with planets. This term was coined by the astronomer William Herschel in the eighteenth century. When viewed through the telescopes of the day, these glowing objects, often green in appearance, looked like planets similar to Uranus. Astronomers have continued to use this term, but do not be fooled—we now know that planetary nebulae indicate the presence of dying stars, not planets.

Focus Question 12-1
In which luminosity class will the Sun be just before it becomes a planetary nebula?

1 Planetary nebulae are quite common in our Galaxy. More than 3000 have been identified, and astronomers estimate that 20,000 to 100,000 planetary nebulae exist in the Milky Way. Indeed, the fate of the Sun is to shed its outer layers as a planetary nebula. The rate at which the gases are removed in a planetary nebula is so slow that they really are not explosive, compared either to our common conception of an explosion on Earth or to the supernova explosions that we will examine shortly.

Because stellar winds from red giants and subsequent planetary nebulae are so plentiful, astronomers estimate that they return a total of about $5 M_{\odot}$ per year of gas and dust to the disk of our Galaxy. This mass amounts to about 85% of all matter expelled by all types of stars. This material goes into the formation of new, metal-rich Population I stars, along with their associated planets and other orbiting bodies. Thus, planetary nebulae play an important role in the chemical and physical evolution of the Galaxy.

The outflowing gases ejected in a planetary nebula take a breathtaking variety of shapes when they interact with gases surrounding their stars, with companion stars, and with the stars' magnetic fields, which are often 10 to 100 times stronger than the Sun's field (Figure 12-4 and Figure 12-5). The Hourglass Nebula (Figure 12-5c) appears initially to have shed mass in a doughnut shape around itself. In the star's final death throes, this gas and dust forced the outflow to go in two directions perpendicular to the plane of the doughnut, creating what is called a *bipolar planetary nebula*.

Spectroscopic observations of planetary nebulae show bright emission lines of hydrogen, carbon, neon, magnesium, oxygen, and nitrogen. From the Doppler shifts of these lines, we conclude that the gas and dust are moving outward from dying intermediate-mass stars at speeds of 10 to 30 km/s (6.2 to 18.6 mi/s). A typical planetary nebula moving outward for 10,000 years has a diameter of a few light-years.

By astronomical standards, a planetary nebula is short-lived. After about 50,000 years, its gases have spread over distances so far from the cooling central star that its nebulosity simply fades from view. The gases then mingle and mix with the surrounding interstellar medium, enriching it with some relatively low-mass elements, such as helium, carbon, oxygen, and nitrogen.

12-2 The burned-out core of an intermediate-mass star becomes a white dwarf

The exposed carbon-oxygen cores of intermediate-mass stars are stable objects supported by electron degeneracy pressure. Such burned-out hulks are called **white dwarfs**, and this is the fate of the Sun's core. Typical white dwarfs are roughly the size of Earth and are covered with a thin coating of hydrogen and helium. At-

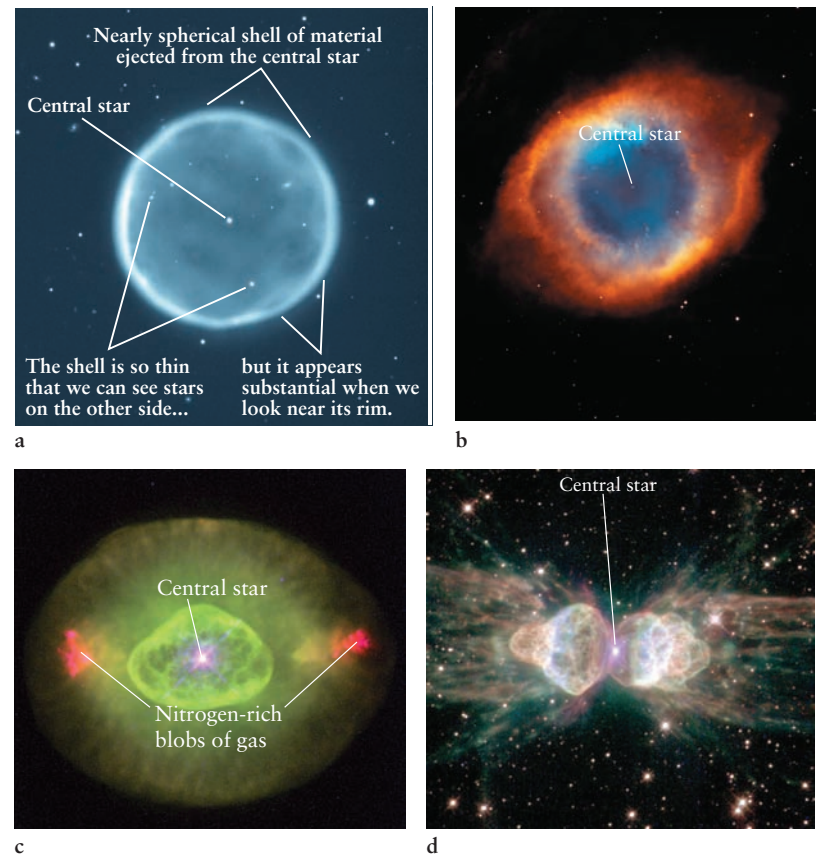


FIGURE 12-4 Some Shapes of Planetary Nebulae The outer shells of dying intermediate-mass stars are ejected in a wonderful variety of patterns. (a) An exceptionally spherical remnant, this shell of expanding gas, in the globular cluster M15 in the constellation Pegasus, is about 7000 ly (2150 pc) away from Earth. (b) The Helix Nebula, NGC 7293, located in the constellation Aquarius, about 700 ly (215 pc) from Earth, has an angular diameter equal to about half that of the full Moon. Red gas is mostly hydrogen and nitrogen, whereas the blue gas is rich in oxygen. (c) NGC 6826 shows, among other features, lobes of nitrogen-rich gas (red). The process by which they were ejected is as yet unknown. This planetary nebula is located in Cygnus. (d) Mz 3 (Menzel 3), in the constellation Norma (the Carpenter's Square), is 3000 ly (900 pc) from Earth. The dying star, creating these bubbles of gas, may be part of a binary system. (a: WIYN/NOAO/NSF; b: NASA, NOAO, ESA, the Hubble Helix Nebula Team, M. Meixner [STScI], and T. A. Rector [NRAO]; c: Bruce Balick [University of Washington], Jason Alexander [University of Washington], Arsen Hajian [U.S. Naval Observatory], Yervant Terzian [Cornell University], Mario Perinotto [University of Florence, Italy], Patrizio Patriarchi [Arcetri Observatory, Italy] and NASA; d: NASA, ESA, and the Hubble Heritage Team [STScI/AURA]. Acknowledgment: R. Sahai [Jet Propulsion Lab] and B. Balick [University of Washington])

mospheres also surround them. Spectra of several white dwarf atmospheres contain elements consistent with Earthlike planets having been vaporized in them. The densities of white dwarfs are typically 10^9 kg/m^3 . In other words, a teaspoonful of white dwarf matter brought to Earth would weigh 5 tons.

Theoretical evolutionary tracks of three burned-out stellar cores are shown in Figure 12-3. These particular white dwarfs evolved from main-sequence stars

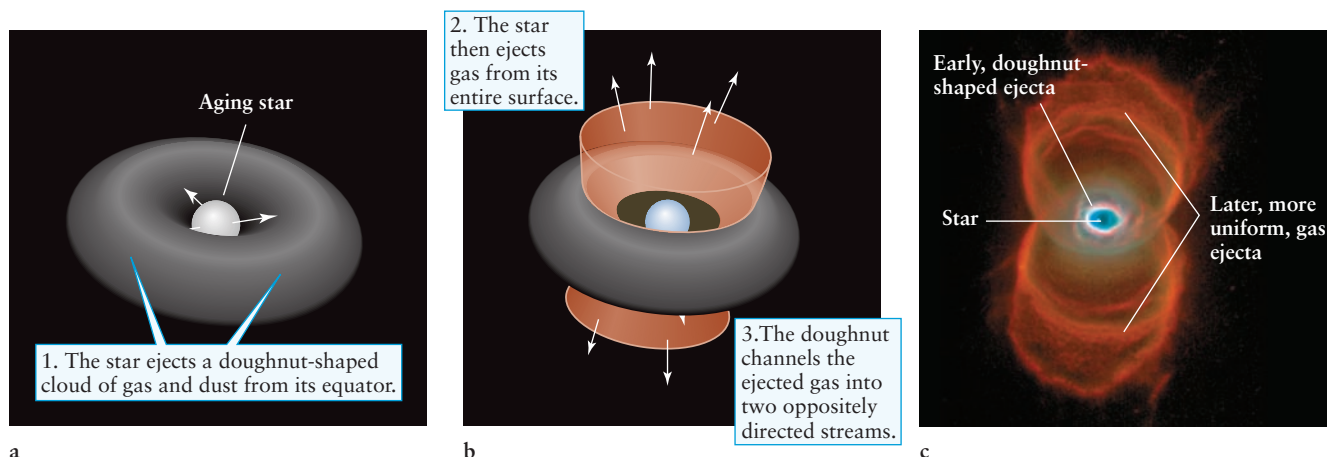


FIGURE 12-5 Formation of a Bipolar Planetary Nebula Bipolar planetary nebulae may form in two steps. Astronomers hypothesize that (a) first, a doughnut-shaped cloud of gas and dust is emitted from the star's equator, (b) followed by outflow that is channeled by the original gas to squirt out perpendicularly to the plane of the doughnut. (c) The

Hourglass Nebula appears to be a “textbook” example of such a system. The bright ring is believed to be the doughnut-shaped region of gas lit by energy from the planetary nebula. The Hourglass is located about 8000 ly (2500 pc) from Earth. (c: Raghvendra Sahai and John Trauger [JPL], the WFPC2 science team, and NASA)

of between $0.8 M_{\odot}$ and $3.0 M_{\odot}$. During the ejection phase, the appearance of these remnants changes rapidly. They appear to race along their evolutionary tracks on the H-R diagram, sometimes executing loops called *thermal pulses* as they temporarily reheat (tracks B and C in Figure 12-3). Finally, as the ejected planetary nebulae fade and the stellar cores cool, the evolutionary tracks of the dying stars take a sharp turn downward toward the white dwarf region of the diagram.

As billions of years pass, an isolated white dwarf radiates its stored energy into space, cooling and decreasing in luminosity. It moves down and to the right on the H-R diagram. The Sun will become a cold, dark, dense sphere of degenerate gases, primarily carbon and oxygen, about the size of Earth. We can actually follow the theory a little further.

Focus Question 12-2

Of what two elements is a white dwarf composed?

Astronomers predict that after billions of years of radiating away their heat, the interior temperatures of white dwarfs will decrease to about 4000 K, and the carbon and oxygen in them will solidify, transforming them into giant crystals.

Many white dwarfs are found in our solar neighborhood, but all are too faint to be seen with the naked eye. The first white dwarf to be discovered is a companion to the bright star Sirius. The binary nature of Sirius was first deduced in 1844 by the German astronomer Friedrich Bessel, who noticed that the star was moving back and forth, as if orbited by an unseen object. This companion, Sirius B (Figure 12-6), was first glimpsed in 1862. Recent satellite observations at ultraviolet wavelengths—where white dwarfs emit most of their light—demonstrate that the surface temperature of Sirius B is about 30,000 K.

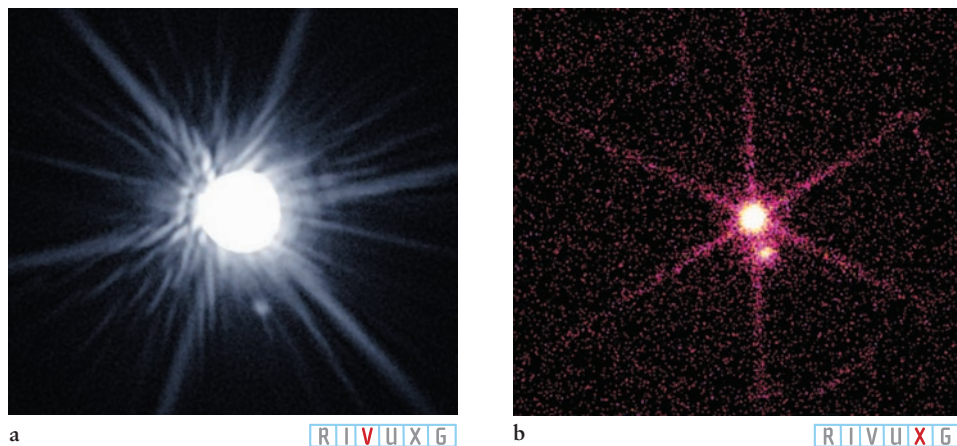
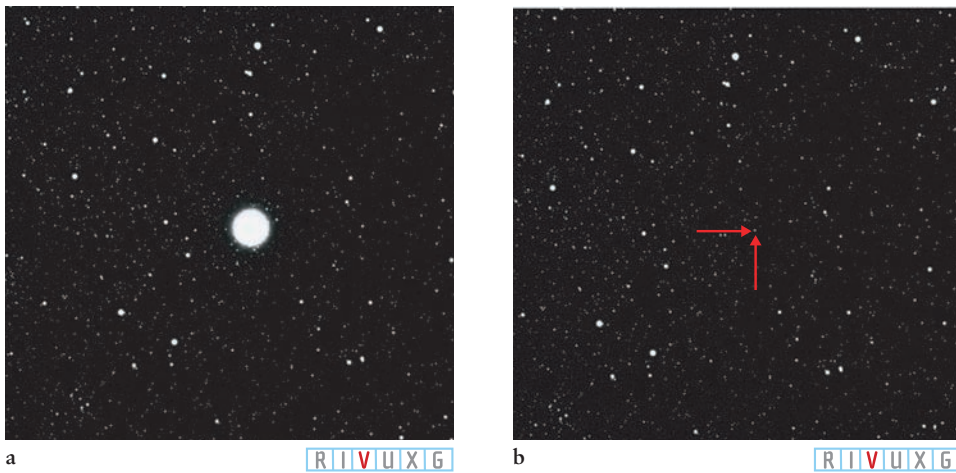


FIGURE 12-6 Sirius and Its White Dwarf Companion (a) Sirius, the brightest-appearing star in the night sky, is actually a double star. The smaller star, Sirius B, is a white dwarf, seen here at the five o'clock position in the glare of Sirius. The spikes and rays around the bright star, Sirius A, are created by optical effects within the telescope. (b) Since Sirius A (11,000 K) and Sirius B (30,000 K) are hot blackbodies, they are strong emitters of X-rays. (a: Damian Peach; b: NASA/SAO/CXC)

**FIGURE 12-7 Nova Herculis 1934**

These two pictures show a nova (a) shortly after peak brightness as a magnitude -3 star and (b) 2 months later, when it had faded to magnitude $+12$. Novae are named after the constellation and year in which they appear. (Lick Observatory/Science Photo Library/Science Source)

White dwarfs in close binary systems can create powerful explosions Occasionally, a star in the sky suddenly becomes between 10^4 and 10^6 times brighter than it is normally. This phenomenon is called a **nova**. Its abrupt rise in brightness is followed by a gradual decline that may stretch over several months or more (Figures 12-7 and 12-8). Astronomers typically observe two or three novae in our Galaxy each year. Novae are fairly common, with about 20 estimated to occur in the Milky Way Galaxy annually. We do not see all of them because of the interstellar gas and dust that obscures many parts of the Galaxy.

2 Painstaking observations strongly suggest that novae occur in close binary systems that contain a white dwarf. The ordinary companion star fills its Roche lobe (recall Figures 11-30, 11-31, and 11-32), so it gradually deposits fresh hydrogen onto the white dwarf. This new mass becomes a dense layer covering the hot surface of the

white dwarf. As more gas is deposited and compressed, the temperature in the hydrogen layer continues to increase. Finally, at about 10^7 K, hydrogen fusion ignites throughout the layer, blowing it into interstellar space. This explosion is the nova. After a nova, fusion ceases on the white dwarf. The companion star, however, may retain enough mass to supply a new layer of surface hydrogen, enabling some novae to reoccur. White dwarfs are not damaged by novae initiated on their surfaces.

Recent visible observations by the Hubble Space Telescope and X-ray observations by the Chandra X-ray Observatory reveal that novae are much more complex than astronomers had thought. Rather than smooth shells of expanding gas, they are composed of thousands of clumps of gas. These observations are stimulating astrophysicists to reexamine the details of the nova mechanism described above to include these newly observed effects.

Theoretical models predict that degenerate matter, such as the electrons in a white dwarf, can withstand only a finite amount of pressure, above which the degenerate matter cannot exert enough outward force to prevent the core from collapsing further. This limits the mass of a white dwarf to less than $1.4 M_{\odot}$. This mass is called the **Chandrasekhar limit**, after the astrophysicist Subrahmanyan Chandrasekhar, who received a Nobel Prize in 1983 for his pioneering theoretical studies of the structure of stars a half-century earlier.

Focus Question 12-3

Why are novae not associated with isolated white dwarfs?

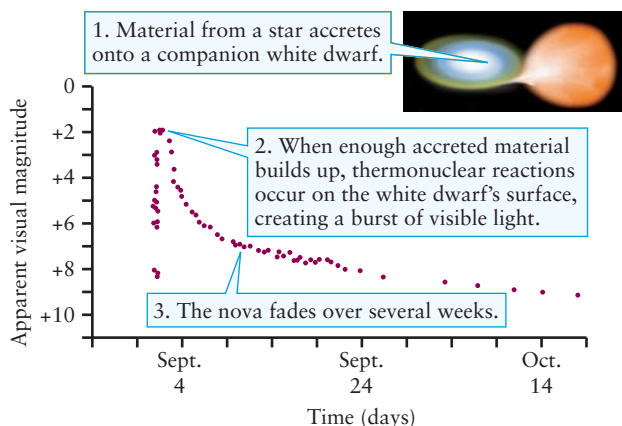


FIGURE 12-8 The Light Curve of a Nova This graph shows the history of Nova Cygni 1975, a nova that was observed to blaze forth in the constellation of Cygnus in September 1975. The rapid rise in magnitude followed by a gradual decline is characteristic of many novae, although some oscillate in intensity as they become dimmer.

Accreting white dwarfs in close binary systems can also explode as Type Ia supernovae Whereas a nova only removes hydrogen and helium that build up on the surface of a white dwarf, some of these stellar remnants actually blow apart completely. This event is called a **Type Ia supernova**, and it occurs in white dwarfs that are part of semidetached binary systems. To trigger most Type Ia supernovae, a companion star dumps gas onto a white dwarf that is close to the Chandrasekhar

limit. When the added gas causes the white dwarf's mass to cross that limit, the increased pressure deep inside the dwarf enables carbon fusion to begin in the core. In a catastrophic runaway process reminiscent of the helium flash, the rate of carbon fusion skyrockets, and with no outer layers to absorb the energy, the star blows up.

In most cases, the companion star survives the massive explosion to some degree. If the companion were a giant or supergiant, the supernova would make the gas in the companion's outer layers glow much more brightly than is actually seen around young Type Ia supernova remnants. Therefore, astronomers believe that the companion stars are main-sequence stars similar to our Sun, because their afterglows from the supernova would be much dimmer than would occur if they were giants or supergiants. In 2012, astronomers reported observing a Type Ia supernova with no afterglow at all from a companion, leading to the deduction that some of these supernovae occur when two white dwarfs in a binary system spiral together and either collide or have the more massive white dwarf pull off the outer layers of the less massive one. In either case, the mass exceeds the Chandrasekhar limit and the white dwarf explodes.

A Type Ia supernova is powered by nuclear energy. What we see is the fallout from a gigantic thermonuclear explosion, which produces a wide array of elements, including silicon, calcium, sulfur, iron, cobalt, and nickel. Especially abundant is a radioactive (unstable) isotope of nickel that decays into a radioactive isotope of cobalt. Most of the electromagnetic display of a Type Ia supernova results directly from the radioactive decay of the nickel and cobalt.

A Type Ia supernova's spectrum lacks hydrogen lines because the white dwarf from which it came was virtually free of this element. The light curve of the supernova, which shows its change of brightness over time, begins with a sudden rise (Figure 12-9). The supernova typically reaches an absolute magnitude of -19 at peak brightness and then declines gradually for more than a year.

Because these supernovae are so bright, they can be seen far beyond the bounds of the Milky Way. Whereas our Galaxy is about 10^5 ly in diameter, Type Ia supernovae can be seen in other galaxies at distances of over 10^9 ly away. Because they all peak at the same absolute magnitude, these supernovae are used to determine distances to myriad galaxies throughout the universe. This calculation is done in two steps. First, the apparent magnitude of a Type Ia supernova at its peak brightness is observed. Second, because we know its absolute magnitude, we use the distance–magnitude relationship (see Discovery 10-2: The Distance–Magnitude Relationship) to calculate its distance.

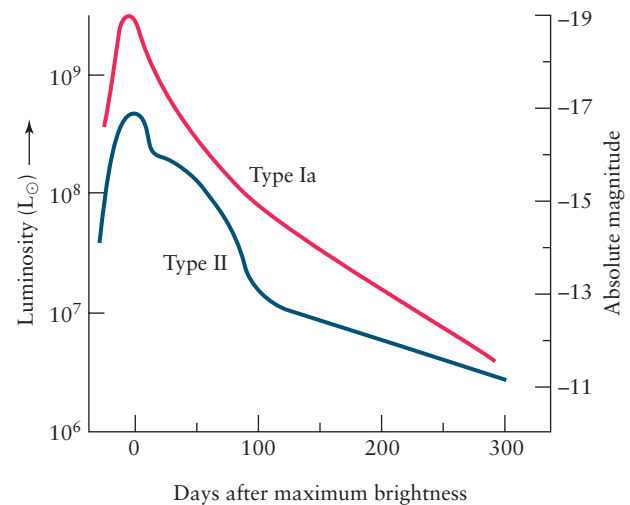


FIGURE 12-9 Supernova Light Curves A Type Ia supernova, which gradually declines in brightness, is caused by an exploding white dwarf in a close binary system. A Type II supernova is caused by the explosive death of a massive star and usually has alternating intervals of steep and gradual declines in brightness.

HIGH-MASS STARS (GREATER THAN $8 M_{\odot}$) AND TYPE II SUPERNOVAE

Due to their increased gravitationally driven compression, stars with more than $8 M_{\odot}$ heat their cores enough to cause the carbon and oxygen there to fuse. As a result, their evolution and fate are very different from those of intermediate-mass stars.

12-3 A series of fusion reactions in high-mass stars leads to luminous supergiants

When helium fusion ends in the core of a star more massive than $8 M_{\odot}$, the inward pressure created by the gravitational force from its outer layers compresses the carbon-oxygen core, driving the star's central temperature above 600 million K. Now carbon fusion begins, producing such elements as neon and magnesium. When a star compresses itself enough to elevate its central temperature to 1.2 billion K, neon fusion occurs, creating oxygen and more magnesium. When the central temperature of the star then reaches 1.5 billion K, oxygen fusion begins, producing sulfur, silicon, and phosphorus. The process of converting lower-mass elements into higher-mass ones is called **nucleosynthesis**.

As the core temperature increases, the core's fusion rate soars. Also, each succeeding core stage has fewer and fewer nuclei than the previous one. For example,

because each helium atom is made up of four hydrogen atoms, only one-quarter as many helium atoms exist in a star's core as there were hydrogen atoms. Combining the higher fusion rate and fewer fusible atoms causes each cycle of core thermonuclear fusion to occur more rapidly than the previous stage. For example, detailed calculations for a star that was initially $25 M_{\odot}$ on the main sequence reveal that carbon fusion occurs for 600 years, neon fusion for 1 year, and oxygen fusion for only 6 months. After half a year of core oxygen fusion, gravitational compression forces the central temperature up to 2.7 billion K, and silicon fusion begins. This thermonuclear process proceeds so furiously that the entire core supply of silicon in a $25 M_{\odot}$ star is converted into iron in 1 day.

Each stage of fusion adds a new shell of matter outside the core (Figure 12-10), creating a structure that resembles the layers of an enormous onion. Together, the tremendous numbers of fusion-generated photons created in all of these shells, as well as in the core, push the outer layers of the star farther and farther outward. The star expands to become a luminous supergiant, almost as wide across as Jupiter's orbit around the Sun.

High-mass main-sequence stars evolve into luminous supergiant stars that are brighter than $10^5 L_{\odot}$, emit winds throughout most of their existence, and have mass-loss rates that exceed those of giants (see Section 11-9). Figure 12-11 shows a supergiant star losing mass. Betelgeuse (see Figure 10-2a), 470 ly away in the constellation of Orion, is another good example of a supergiant experiencing mass loss. Spectroscopic observations show that Betelgeuse is losing mass at the rate of $1.7 \times 10^{-7} M_{\odot}$ per year and is surrounded by ejected gas in a *circumstellar shell*. This huge shell is expanding at 10 km/s (6.2 mi/s), and escaping gases have been de-

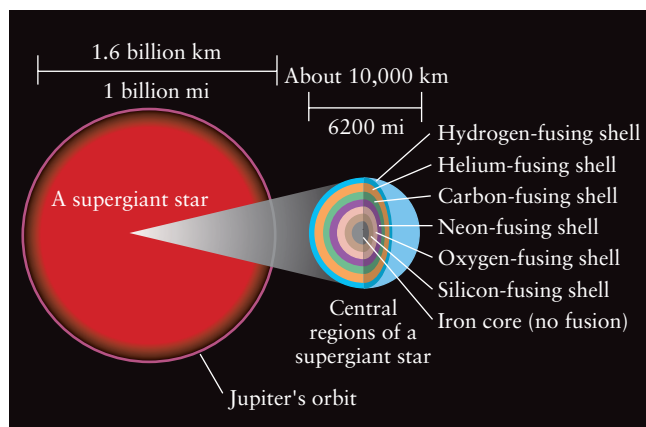


FIGURE 12-10 The Structure of an Old High-Mass Star Near the end of its life, a high-mass star becomes a supergiant with a diameter almost as wide as the orbit of Jupiter. The star's energy comes from six concentric fusing shells, all contained within a volume roughly the same size as Earth.



FIGURE 12-11 Mass Loss by a Supermassive Star Gas and dust ejected by the massive star HD 148937 in the constellation Norma. Orange is hydrogen, blue is oxygen, and red is sulfur. Located about 4200 ly (1300 pc) away, this star has $40 M_{\odot}$ and is more than halfway through its 6-million-year life span. (Gemini Observatory/AURA)

tected at distances of 10,000 AU from the star. The expanding circumstellar shell has an overall diameter of one-third of a light-year.

Silicon fusion in high-mass stars involves many types of nuclear reactions, but its major final product is iron. Eventually, the core is converted entirely into iron, which is surrounded by layers of shell fusion that consume the remaining reserves of fuel deep inside the star (see Figure 12-10). While the star's enormously bloated atmosphere is nearly as big as the orbit of Jupiter, its entire energy-producing region is contained in a volume the size of Earth. The buildup of an inert core of iron nuclei and electrons signals the impending violent death of a massive star.

12-4 High-mass stars blow apart in Type II supernova explosions

Unlike lighter elements, iron cannot fuel further thermonuclear reactions. The protons and neutrons inside iron nuclei are already so tightly bound together that no further energy can be extracted by fusing still more particles with them. The sequence of fusion stages in the cores of high-mass stars therefore ends.

Focus Question 12-4

Where in massive stars are the higher-mass elements formed?

TABLE 12-1 EVOLUTIONARY STAGES OF A 25 M_☉ STAR

Stage	Central temperature (K)	Central density		Duration of stage
		(kg/m ³)	(lb/yd ³)	
Hydrogen fusion	4 × 10 ⁷	5 × 10 ³	8.4 × 10 ³	7 × 10 ⁶ years
Helium fusion	2 × 10 ⁸	7 × 10 ⁵	1.2 × 10 ⁶	5 × 10 ⁵ years
Carbon fusion	6 × 10 ⁸	2 × 10 ⁸	3.4 × 10 ⁸	600 years
Neon fusion	1.2 × 10 ⁹	4 × 10 ⁹	6.7 × 10 ⁹	1 year
Oxygen fusion	1.5 × 10 ⁹	1 × 10 ¹⁰	1.7 × 10 ¹⁰	6 months
Silicon fusion	2.7 × 10 ⁹	3 × 10 ¹⁰	5.1 × 10 ¹⁰	1 day
Core collapse	5.4 × 10 ⁹	3 × 10 ¹²	5.1 × 10 ¹²	0.2 second
Core bounce	2.3 × 10 ¹⁰	4 × 10 ¹⁷	6.7 × 10 ¹⁷	milliseconds
Supernova explosion	about 10 ⁹	varies	varies	hours

Because iron atoms do not fuse and emit energy, the electrons in the core must now support the star’s outer layers by the strength of electron degeneracy pressure alone. Soon, however, the continued deposition of fresh iron from the silicon-fusing shell causes the core’s mass to exceed the Chandrasekhar limit. Electron degeneracy suddenly fails to support the star’s enormous weight, and the core collapses.

Any isolated main-sequence star of more than 8 M_☉ develops an iron core as a luminous supergiant. When the core collapses, a rapid series of cataclysms is triggered that tears the star apart in a few seconds. Let us see how this happens in the death of a 25 M_☉ star, according to computer simulations (Table 12-1).

In the 25 M_☉ star, electron degeneracy pressure fails when the density is sufficiently high inside the iron core. The core, only some 3000 km in diameter, then collapses immediately. In roughly $\frac{1}{10}$ second, the central temperature exceeds 5 billion K. Gamma-ray photons associated with this intense heat have so much energy that they begin to break apart the iron nuclei, a process called **photodisintegration**. Although millions of years passed from the time the star arrived on the main sequence with a hydrogen- and helium-filled core until its core became iron, it takes less than a second to convert the core back into elemental protons, neutrons, and electrons!

Within another $\frac{1}{10}$ second, as the density continues to climb, the electrons in the core are forced to combine with the core’s protons to produce neutrons, and the process releases a flood of neutrinos. As we saw in Section 9-10, neutrinos have no electric charge and have very little mass, and most pass through Earth or the Sun without interaction. Nevertheless, the matter deep inside a collapsing high-mass star is so fantastically dense that its newly created neutrinos cannot immediately escape from the star’s core. Slamming into nearby particles, the neutrinos provide a pressure pushing outward from the core.

According to the calculations, about $\frac{1}{4}$ second after the collapse begins, the density of the entire core reaches 4 × 10¹⁷ kg/m³ (6.7 × 10¹⁷ lb/yd³), which is *nuclear density*, the density at which neutrons and protons are normally packed together inside individual atomic nuclei. (Compare this with the density of water, 10³ kg/m³ (1700 lb/yd³).) When the neutron-rich material of the core reaches nuclear density, it suddenly stiffens. As a result, the core resists the collapse much more than when it was less dense. The collapse of the core halts so abruptly that it rebounds and begins rushing back out in a process called *core bounce*.

The model predicts that, during this critical stage, the star’s unsupported layers of shell-fusing matter outside the core are plunging inward at up to 15% of the speed of light. The outward-flowing neutrinos and rebounding core slam into this matter (the red regions in Figure 12-12a). The impact stops the core’s rebound while causing the infalling matter to reverse course. In just a fraction of a second, a tremendous volume of matter begins to move back up toward the star’s surface. This matter accelerates rapidly as it encounters less and less resistance, and soon it forms an outgoing shock wave, analogous to the thunder created by a lightning flash. After a few hours, this shock wave reaches the star’s surface, lifting the star’s outer layers away from the core in a colossal blast. The star becomes a **Type II supernova**.

As the supernova expands, the star’s luminosity, which was already 10⁴ L_☉ to 10⁵ L_☉, suddenly increases by another factor of between 10³ and 10⁶. A supernova is often 10% as bright as all of the other stars in its galaxy combined. The brightest supernova that has been observed emitted as much energy as a hundred billion stars as luminous as our Sun!

3 Many of the elements with more protons than iron are created during supernovae. In the twentieth century, physicists discovered that neutrons can transform into protons by emitting an electron. As the layers of a

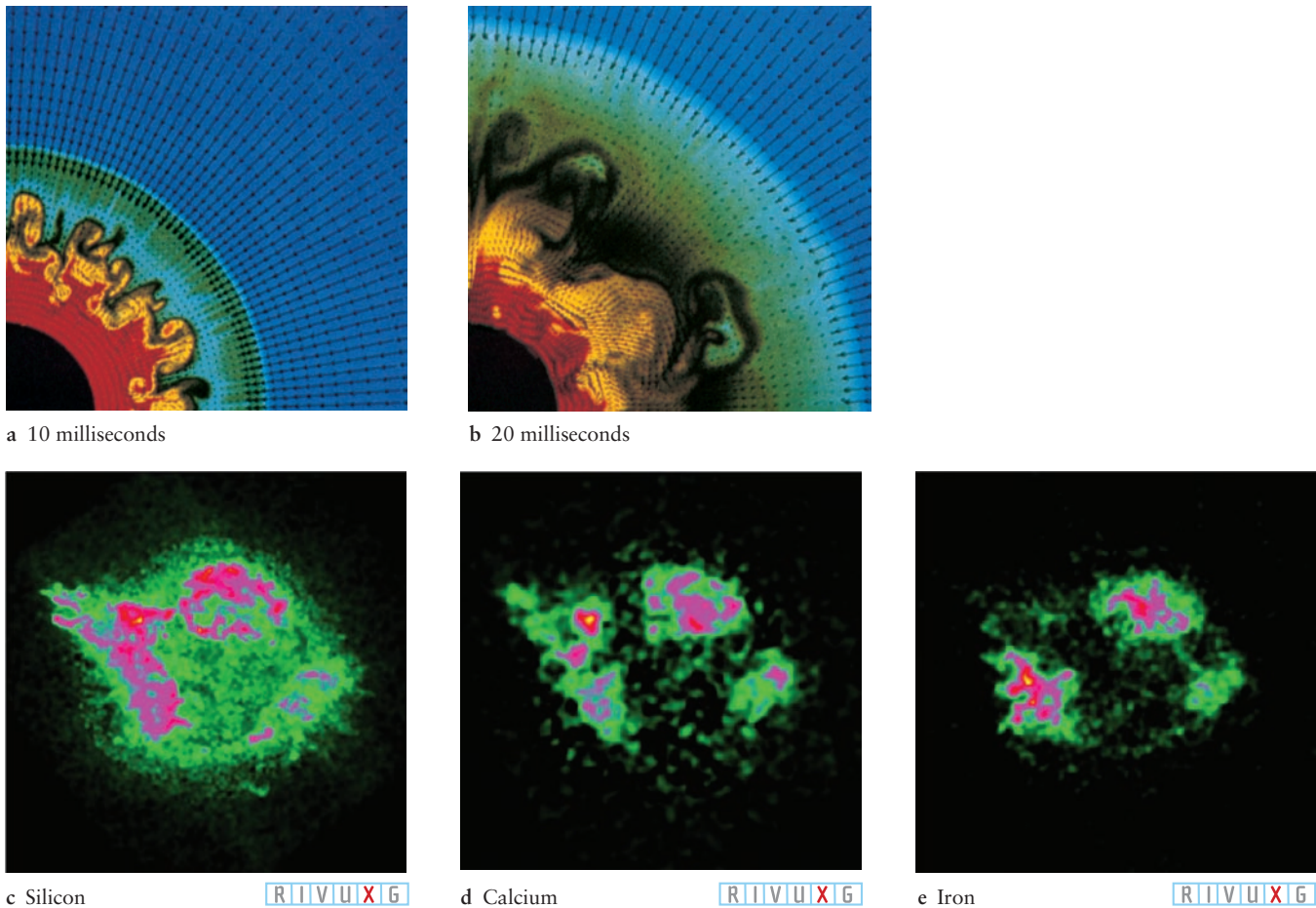


FIGURE 12-12 Supernovae Proceed Irregularly Images (a) and (b) are computer simulations showing the chaotic flow of gas deep inside the star as it begins to explode as a supernova. This uneven flow helps account for the globs of iron and other heavy elements emitted from deep inside, as well as the lopsided distribution

of all elements in the supernova remnant, as shown in (c), (d), and (e). These three pictures are X-ray images of supernova remnant Cassiopeia A taken by the Chandra Observatory at different wavelengths. (a–b: Courtesy of Adam Burrows, University of Arizona, and Bruce Fryxell, NASA/GSFC; c–e: U. Hwang et al./GSFC/NASA)

massive dying star are blasted into space, they are compressed so much by the neutrinos and the shock wave that neutrons are forced into the elements that have been created as the star evolved. Packed with extra neutrons, many of these nuclei are unstable. As a result, neutrons begin emitting electrons, a process that converts these neutrons to protons. Once this occurs, that atomic nucleus is no longer the element it was a moment before. For example, a neutron-enriched iron atom transforms into cobalt when a neutron transforms to a proton in it. Whereas the neutron-enriched iron was not stable, the cobalt it becomes can be—an atom of cobalt has been created. This process creates many of the heavy elements that exist today. Likewise, the supernova can cause atoms to bond together and create interstellar dust particles (see Sections 4-1 and 4-2). In summary, the products of shell fusion and this final burst of fusion during supernovae are where many of the heavy elements on Earth and other terrestrial planets come from, including some of the atoms in our bodies. We really are made of stardust.

Computer simulations indicate that the onion skins of different elements do not just move out of a star going supernova in lockstep (Figure 12-12a, b). Therefore, the ejected gases are not expected to be uniform shells of matter. Indeed, observations reveal that the material emerging in a supernova comes out quite irregularly. Figure 12-12c–e shows the distributions of some of the elements in the supernova Cassiopeia A (Cas A). Other observations show that some of the dense elements, like iron, ejected in supernovae are actually created deep inside the innermost silicon and oxygen layers of the supernovae and then shot out like bullets or geysers. Other elements are created throughout the expanding stars. However, calculations and observations reveal that supernovae do not create enough of the heaviest elements, like gold, silver, and platinum, to account for the amounts of these elements found on Earth. We will explore their likely origin in Section 12-9.

Focus Question 12-5

What role do neutrinos play in Type II supernovae?

Insight Into Science

The Process of Science The best scientific theories and models also provide explanations about matters that they were not initially designed to study. In modeling stellar evolution and supernovae, the processes that create and eject metals (that is, elements heavier than hydrogen and helium) from stars also provide the elements necessary for life to exist.

Focus Question 12-6

Why does most of the iron on Earth not come from the iron cores of high-mass stars that explode as Type II supernovae?

Over its lifetime, our model $25 M_{\odot}$ star ejects more than $20 M_{\odot}$ of its mass back into space—by stellar winds, by gas ejected during unstable periods, and during its supernova phase. Less massive stars return proportionately less mass to the interstellar medium.

The spectrum and light curve of a Type II supernova are both significantly different from those of a Type Ia supernova. Whereas hydrogen lines are absent from a Type Ia's spectrum, they are strong in the spectra of Type II supernovae. Both types begin with a sudden rise in brightness (see Figure 12-9). However, a Type II supernova usually peaks at $M = -17$, a hundred times dimmer than a Type Ia. Whereas Type Ia supernovae decline gradually for more than a year, Type II supernovae alternate between periods of steep and gradual declines in brightness. Type II light curves, therefore, have a steplike appearance.

12-5 Supernova remnants are observed in many places

Hundreds of supernovae are observed each year. For example, 232 were seen in 2013. Several thousand have been studied and catalogued. These observations suggest that in a typical galaxy like the Milky Way, Type Ia supernovae occur approximately once every 36 years, whereas Type II supernovae occur about once every 44 years. Thus, there should be about five supernovae exploding in our Milky Way Galaxy each century. Vigorous stellar evolution in our Milky Way occurs primarily in the Galaxy's disk, because stars there continue to form in its giant molecular clouds. The disk (which we are in and which we see edge-on as the milky white band of stars and glowing gas and dust across the sky) is also where supernovae explode.

Astronomers find the debris of local supernova explosions in the Milky Way scattered across the Galaxy's disk. A beautiful example is the Cygnus Loop (see Figure 11-5). The star's outer layers were blasted into space 20,000 years ago so violently that they are still traveling at supersonic speeds. As this expanding shell of gas plows through the interstellar medium, it collides with atoms and molecules, making the gases glow.

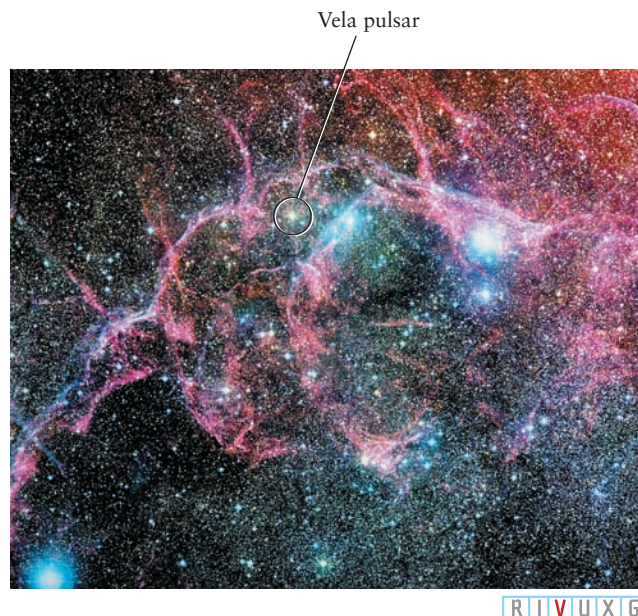


FIGURE 12-13 The Gum Nebula The Gum Nebula is the largest known supernova remnant. It spans 60° across the sky and is centered roughly on the southern constellation of Vela. The nearest portions of this expanding nebula are only 300 ly from Earth. The supernova explosion occurred about 11,000 years ago, and the remnant now has a diameter of about 2300 ly. Only the central regions of the nebula are shown here. (Royal Observatory, Edinburgh/Anglo-Australian Telescope Board/Science Source)

Many supernova remnants cover sizable fractions of the sky. The largest is the Gum Nebula, with an angular diameter of 60° (its central region is shown in Figure 12-13). This nebula looks so big because it is so close: Its near side is only about 300 ly from Earth. Studies of the Gum Nebula's expansion rate suggest that the supernova exploded about 11,000 years ago. At maximum brilliance, the exploding star was probably as bright as the Moon at first quarter.

The Milky Way's disk is so filled with interstellar gas and dust that we cannot see very far into it. Supernovae are thought to erupt every few decades in remote parts of our Galaxy, but their detonations are hidden from optical telescopes by the interstellar medium. Many supernova remnants in our Galaxy can be detected only at nonvisible wavelengths, ranging from X-rays through radio waves. For example, Figure 12-14 shows images of the supernova remnant Cassiopeia A as seen in radio (Figure 12-14a) and X-ray (Figure 12-14b). Visible-light photographs of this part of the sky reveal only a few small, faint wisps. Thus, radio searches for supernova remnants are more fruitful than visual searches. Only 24 supernova remnants have been found in visible-light photographs, but more than 3100 remnants in our Galaxy and in others have been discovered by radio astronomers.

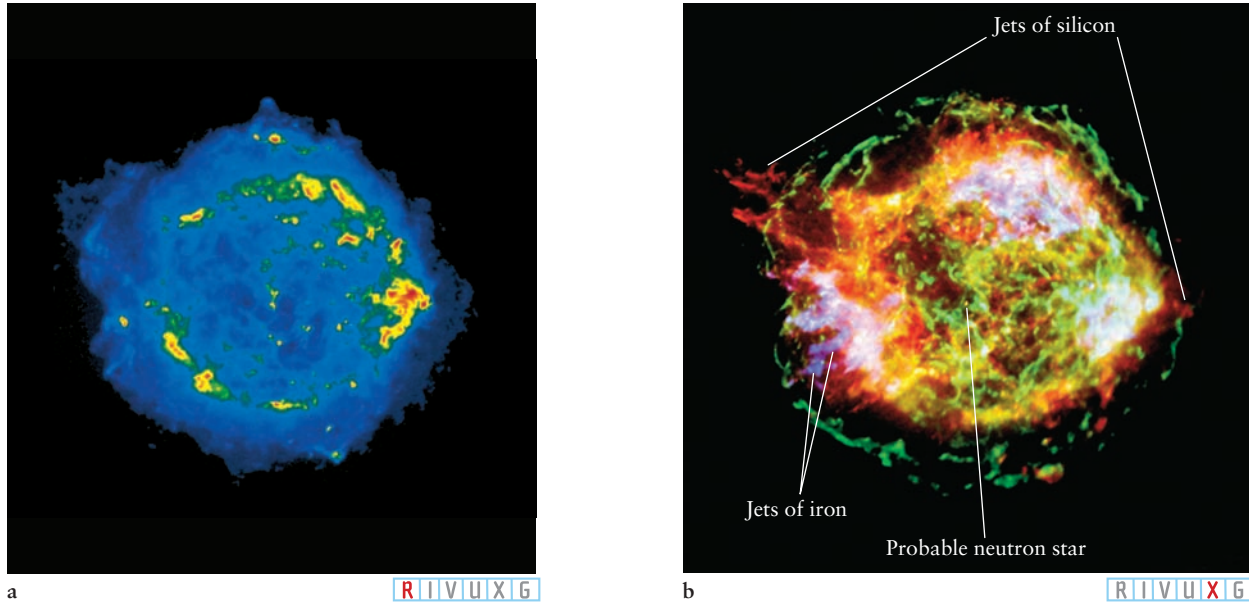


FIGURE 12-14 Cassiopeia A Supernova remnants, such as Cassiopeia A, are typically strong sources of radio waves and X-rays. **(a)** A radio image produced by the Very Large Array (VLA). **(b)** A corresponding X-ray picture of Cassiopeia A taken by the Chandra Observatory. The opposing jets of silicon, probably guided by

powerful magnetic fields, were ejected early in the supernova, before the iron-rich jets were released. Radiation from the supernova that produced this nebula first reached Earth 300 years ago. The explosion occurred about 11,000 ly from Earth. (a: NRAO; b: U. Hwang [GSFC/UMD], J. M. Lamming [NRL], et al., CXC, NASA)

From the expansion rate of the nebulosity in Cassiopeia A, astronomers calculate that the light from that supernova explosion first reached Earth about 300 years ago. Although telescopes were in wide use by the late 1600s, no record of the event is known. In fact, the last supernova of a star near its maximum brightness in our Galaxy, seen by the naked eye, was observed by Johannes Kepler in 1604. In 1572, Tycho Brahe also recorded the sudden appearance of an exceptionally bright star in the sky. While it has been centuries since the last supernova visible to the naked eye occurred in our Galaxy, supernovae are so bright that some naked-eye events in other galaxies have been seen. Perhaps the most notable one is SN 1987a, which we explore in the next section.

In 2002, astronomers discovered that 2 million years ago a group of bright O and B stars, called the Scorpius-Centaurus OB association, passed within 130 ly of Earth, and that one or more supernovae in this group may have occurred at that time. Such a close explosion could have damaged Earth's ozone layer and caused the extinction of some ocean life shortly after the beginning of the Pleistocene epoch on Earth.

Besides the gas and dust enriched with heavy elements, supernovae leave a heritage long after their remnants have dissipated. Much of this gas and dust is recycled into new stars and planets, such as our solar system.

Supernova 1987A offered a detailed look at a massive star's death Some supernovae in other galaxies can be seen without a telescope. In 1885, a supernova in

the Andromeda Galaxy was just barely visible to the unaided eye. On February 23, 1987, a supernova was observed in the Large Magellanic Cloud, a galaxy near our own Milky Way. The supernova, designated SN 1987A, was the first to be observed that year (**Figure 12-15**), and it was so bright that it could easily be seen with the naked eye. What made SN 1987A such an exciting discovery was that it gave astronomers a rare opportunity to study the death of a nearby massive star using modern equipment that provided data with which to verify the theory of supernovae described earlier.

At first, SN 1987A reached only a tenth of the luminosity predicted for an exploding massive star. For the next 85 days, it gradually brightened. Then it began dimming, as is characteristic of an ordinary supernova. Fortunately, the star had been observed before it became a supernova (**Figure 12-15** insets). The Large Magellanic Cloud is about 160,000 ly from Earth—near enough to us that many of its stars have been individually observed and catalogued. The star had been identified as a B3 supergiant. The theory of the evolution of such stars provided an explanation for SN 1987A's unusually slow brightening.

When this star was on the main sequence, its mass was about $20 M_{\odot}$, although by the time it exploded, it had shed many solar masses. The evolutionary track for an aging $20 M_{\odot}$ star wanders back and forth across the top of the H-R diagram, so the star alternates between being a hot (blue) supergiant and a cool (red) supergiant.

Focus Question 12-7

Why are we unable to see all the supernovae in our Galaxy?

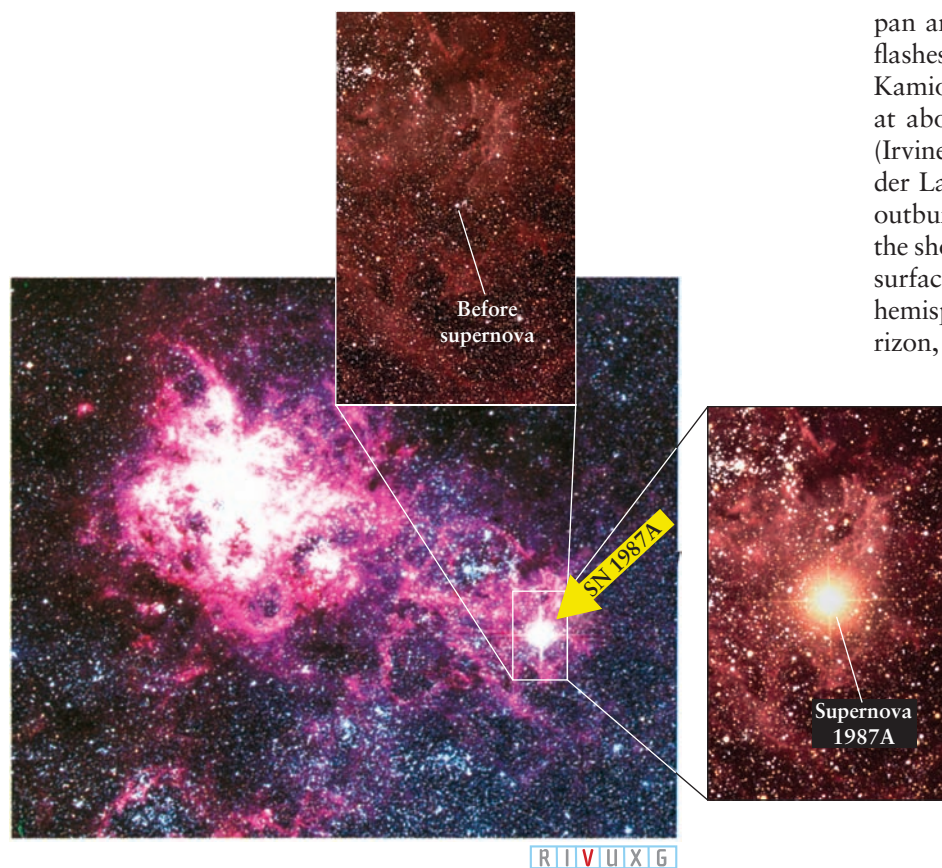


FIGURE 12-15 Supernova 1987A A supernova was discovered in a nearby galaxy called the Large Magellanic Cloud (LMC) in 1987. This photograph shows a portion of the LMC that includes the supernova and a huge H II region, called the Tarantula Nebula. At its maximum brightness, observers at southern latitudes saw the supernova without a telescope. (**Insets**) The star before and after it exploded. (European Southern Observatory; insets: Australian Astronomical Observatory/David Malin Images)

The star's size changes significantly as its surface temperature changes. A blue supergiant is only 10 times larger in diameter than the Sun, but a red supergiant of the same luminosity is 1000 times larger. Because the doomed star was a relatively small blue supergiant when it exploded, it reached only a tenth of the brightness that it would have attained had it been a red supergiant at that time.

Ordinary telescopic observations of a supernova explosion can only show us the expanding outer layers of the dying star. No photons created in the core can penetrate through the supernova gases directly to our telescopes. As the outer layers of the supernova thin, however, neutrinos escape directly from the core into space. By detecting these neutrinos and measuring their properties, astronomers can learn many details about the star's collapsing core, especially about core bounce. Several neutrino detectors were operating when the neutrinos from SN 1987A reached Earth.

Nearly a day before SN 1987A was first observed in the sky, teams of scientists at neutrino detectors in Ja-

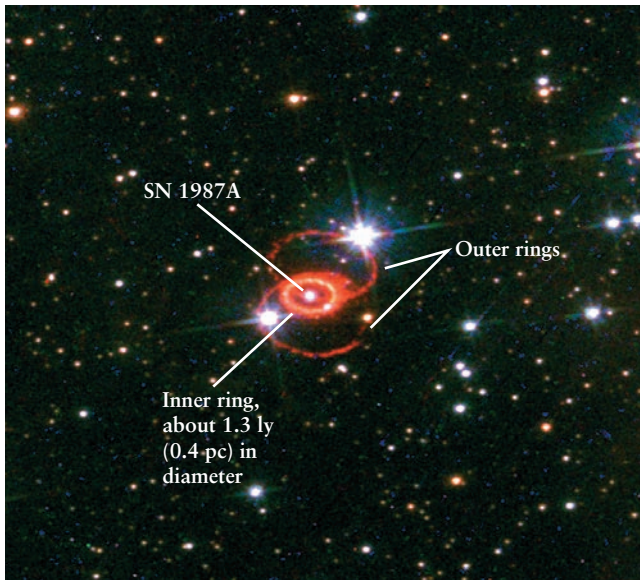
pan and the United States reported observing Cerenkov flashes (see Section 9-10) from a burst of neutrinos. The Kamiokande II detector in Japan detected 12 neutrinos at about the same time that 8 were found by the IMB (Irvine-Michigan-Brookhaven) detector in a salt mine under Lake Erie. Neutrinos preceded the visible supernova outburst because they escaped from the dying star before the shock wave from the collapsing core reached the star's surface. The particles were detected in Earth's northern hemisphere, where the supernova is always below the horizon, after having passed through Earth. The discovery of these neutrinos provided strong support for the theory that supernovae are caused by the collapse of a star's core, its subsequent bounce, and the flow of neutrinos that help cause the supernova, as described in Section 12-6.

About $3\frac{1}{2}$ years after SN 1987A's detonation was first seen from Earth, astronomers using the Hubble Space Telescope obtained a picture showing several rings of glowing gas around the exploded star (Figure 12-16). This gas had been emitted by the star 20,000 years before the supernova, when a hydrogen-rich stellar envelope was ejected by stellar winds from the doomed star, as discussed earlier in this chapter. This gas expanded in an hourglass shape, because it was blocked from expanding around the star's equator, either by a preexisting ring of gas or by the orbit of an as-yet-unseen companion star.

While emitting these gases, the doomed star was a red supergiant. It then shrank and in its final blue giant phase, the same star emitted higher-speed gas that compressed the hourglass gas into three rings—a narrow one around the equator and wider ones at the top and bottom of the hourglass. This relic gas is now being illuminated by photons from the supernova. In 1998, astronomers began to observe gas ejected from the supernova striking the preexisting inner ring of gas. This collision has caused that ring to brighten, and it also illuminates more and more of the gas that had been ejected earlier.

SN 1987A provided astronomers with invaluable information, in part because the doomed star had been studied before it exploded and its distance from Earth was known. The supernova was also located in an unobscured part of the sky, and neutrino detectors happened to be operating at the time of the outburst. Astronomers will be monitoring the progress of this supernova for years to come.

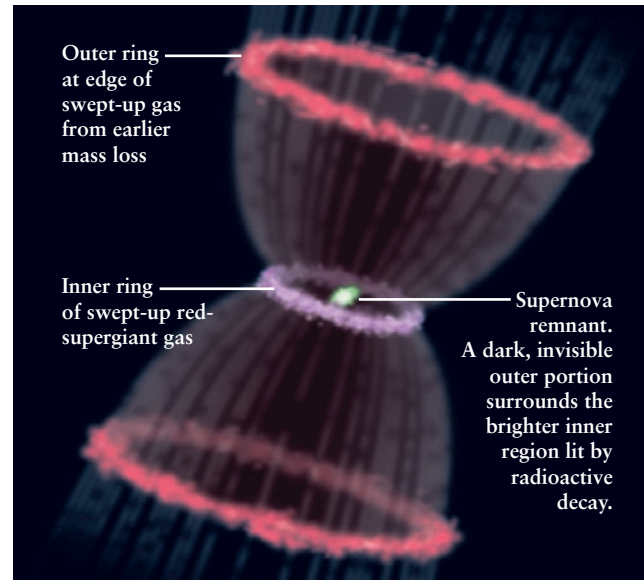
4 Cosmic rays are not rays at all In observations connected with supernovae, astronomers have detected particles flying through space at speeds that exceed 90% of the speed of light. These high-speed particles are called **cosmic rays** or **primary cosmic rays**. They were named *rays* before their true identity as particles was known, and, as often happens, this misleading name has stuck. About 90% of cosmic rays are protons (hydrogen nuclei).



a Supernova 1987A seen in 1996

R I V U X G

FIGURE 12-16 Shells of Gas Around SN 1987A (a) Intense radiation from the supernova explosion caused three rings of gas surrounding SN 1987A to glow in this Hubble Space Telescope image. This gas was ejected from the star 20,000 years before the star detonated. All three rings lie in parallel planes. The inner ring is about 1.3 ly across. The white and colored spots are unrelated stars. (b) When the progenitor star of SN 1987A was still a red supergiant, a slowly moving wind from the star filled the surrounding space with a thin gas. When the star contracted into a blue supergiant, it produced a



b An explanation of the rings

faster-moving stellar wind. The interaction between the fast and slow winds somehow caused gases to pile up along an hourglass-shaped shell surrounding the star. The burst of ultraviolet radiation from the supernova ionized the gas in the rings, causing the rings to glow. The supernova itself, at the center of the hourglass, glows because of energy released from radioactive decay. (a: Robert Kirshner and Peter Challis, Harvard-Smithsonian Center for Astrophysics; STScI; b: Robert Kirshner and Peter Challis, Harvard-Smithsonian Center for Astrophysics; STScI/NASA)

A few percent of cosmic rays are electrons and positrons (positively charged electrons). Less than 1% of cosmic rays are nuclei of more massive, nonhydrogen elements.

Most primary cosmic rays coming in our direction collide with gas in the atmosphere some 15 km (9.3 mi) above Earth's surface. This is fortunate, because each cosmic ray packs an enormous amount of energy that potentially could harm living tissue. A collision in the air divides a cosmic ray's energy among several gas particles, which are shoved Earthward. These particles, in turn, often hit other gas atoms, creating a cascade of lower-energy particles that eventually reaches Earth's surface as a **cosmic ray shower** (Figure 12-17). These cosmic rays created from atoms in our atmosphere are called **secondary cosmic rays**.

The origins of cosmic rays are beginning to become clear. At first, many astronomers thought that moderate-energy cosmic rays were emitted directly by supernovae. This theory made sense because these explosions are the best-known source of the energy necessary to give such cosmic rays their tremendous speeds. However, observations by the *Advanced Composition Explorer* satellite reveal that the isotopes of nickel coming to Earth in the form of cosmic rays are not the isotopes emitted

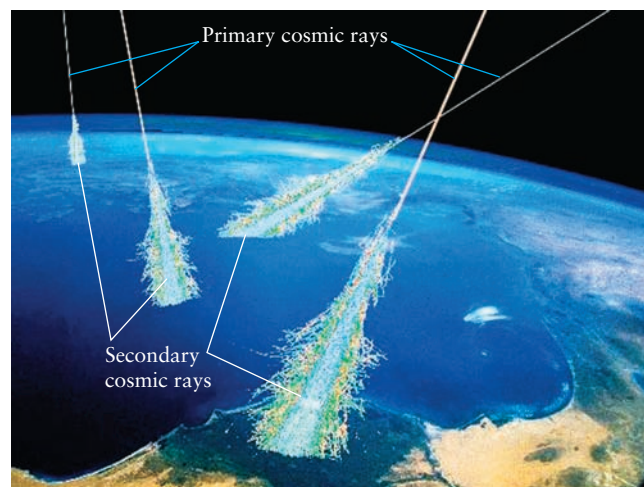


FIGURE 12-17 Cosmic Ray Shower Cosmic rays from space slam into particles in the atmosphere, breaking them up and sending them Earthward. These debris are called secondary cosmic rays. This process of impact and breaking up continues as secondary cosmic rays travel downward, creating a cosmic ray shower, as depicted in this artist's conception of four such events. (Simon Swordy [University of Chicago], NASA)

by supernovae. Instead, evidence collected by the Fermi Gamma-ray Space Telescope in 2013 revealed that many cosmic rays are created when supernova debris slams into gas already in space, causing some of this preexisting matter to accelerate and become cosmic rays.

Besides emitting particles, supernovae create gamma rays when cosmic ray protons created as a result of the supernova smash into normal protons in interstellar space. These collisions create short-lived particles called pions that decay into pairs of gamma rays.

Much less common than cosmic rays from supernova remnants are ultrahigh-energy cosmic rays, some of which come from inside our Galaxy, while many more come from outside our Galaxy. Each of these particles has as much kinetic energy as a baseball thrown by a major league pitcher. In 2004, the Pierre Auger Observatory, consisting of 1600 cosmic ray sensors designed to detect the highest-energy cosmic rays, was commissioned. Located on the Pampa Amarilla (“Yellow Prairie”) in Argentina, it is designed to give astronomers information about the directions from which these cosmic rays come. Observations of their paths suggest that many of them come from galaxies containing supermassive black holes, which we will discuss in Section 12-15. Whether these black holes themselves generate the ultrahigh-energy cosmic rays is still under investigation.

NEUTRON STARS AND PULSARS

When stars that initially had between $8 M_{\odot}$ and $25 M_{\odot}$ explode as supernovae, their expanding cores are stopped by the outer layers into which they slam. The cores recontract and remain intact as highly compressed clumps of neutrons, called **neutron stars**. Isolated neutron stars, resulting from these explosions, are stable stellar remnants with masses between $1.25 M_{\odot}$ and about $3 M_{\odot}$.

12-6 The cores of many Type II supernovae become neutron stars

The neutron was discovered during laboratory experiments in 1932. Inspired by the realization that white dwarfs are supported by electron degeneracy pressure, Fritz Zwicky and Walter Baade proposed that a similar repulsion between neutrons could support remnant neutron cores. They used the fact that the Pauli exclusion principle also prevents neutrons with identical properties from packing too closely together. When neutrons are pressed together, many of them have to move rapidly (so as not to be identical to their neighbors). This motion provides a pressure, called **neutron degeneracy pressure**, that equations predict is even greater than electron degeneracy pressure (see Section 11-10) and thus allows stellar remnants with masses beyond the Chandrasekhar limit to be stable.

Most scientists ignored Zwicky and Baade’s theory for years. After all, a neutron star seemed to be a rather unlikely object. To transform protons and electrons into

neutrons, the density in the star would have to be equal to nuclear density, about 10^{17} kg/m^3 ($1.7 \times 10^{17} \text{ lb/yd}^3$). Thus, a teaspoon full of neutron star matter brought back to Earth would weigh about 1 billion tons. Put another way, neutron star matter is 200 million times denser than matter found in a white dwarf. Furthermore, an object compacted to nuclear density would be very small. A $2 M_{\odot}$ neutron star would have a diameter of only 20 km (12 mi) and would fit inside any major city on Earth. The surface gravity on one of these neutron stars would be so strong that the escape velocity would equal one-half the speed of light. All of these conditions appeared to be so outrageous that few astronomers paid any serious attention to the subject of neutron stars—until 1968.

In that year, Dame Jocelyn Bell Burnell, then just Jocelyn Bell, a graduate student in radio astronomy at Cambridge University, noticed that the radio telescope she was using was detecting regular pulses from one particular location in the sky. Careful repetition of the observations demonstrated that the radio pulses were arriving with a regular period of 1.3373011 seconds. The regularity of this pulsating radio source was so striking that the Cambridge team suspected that they might be detecting signals from an advanced alien civilization, which accounts for the name first assigned to it: LGM1 (short for Little Green Men 1). This possibility was soon discarded as several more of these pulsating radio sources, which came to be known as **pulsars**, were discovered across the sky. The broad distribution of pulsars would require incredibly widespread civilizations, inconsistent with what we have found in our searches for alien intelligence (see Chapter 15). In all cases, the pulsar periods were extremely regular, ranging in those early days between 0.2 and 1.5 seconds (Figure 12-18). The range of known pulsar periods is much broader today—from just over 1.5 milliseconds to just under 9 seconds.

When the discovery of pulsars was officially announced in early 1968, astronomers around the world began proposing all sorts of explanations. Many of these theories were bizarre, and arguments raged for months. We have already ruled out alien civilizations.

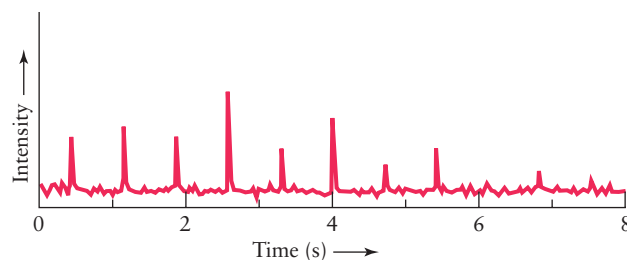
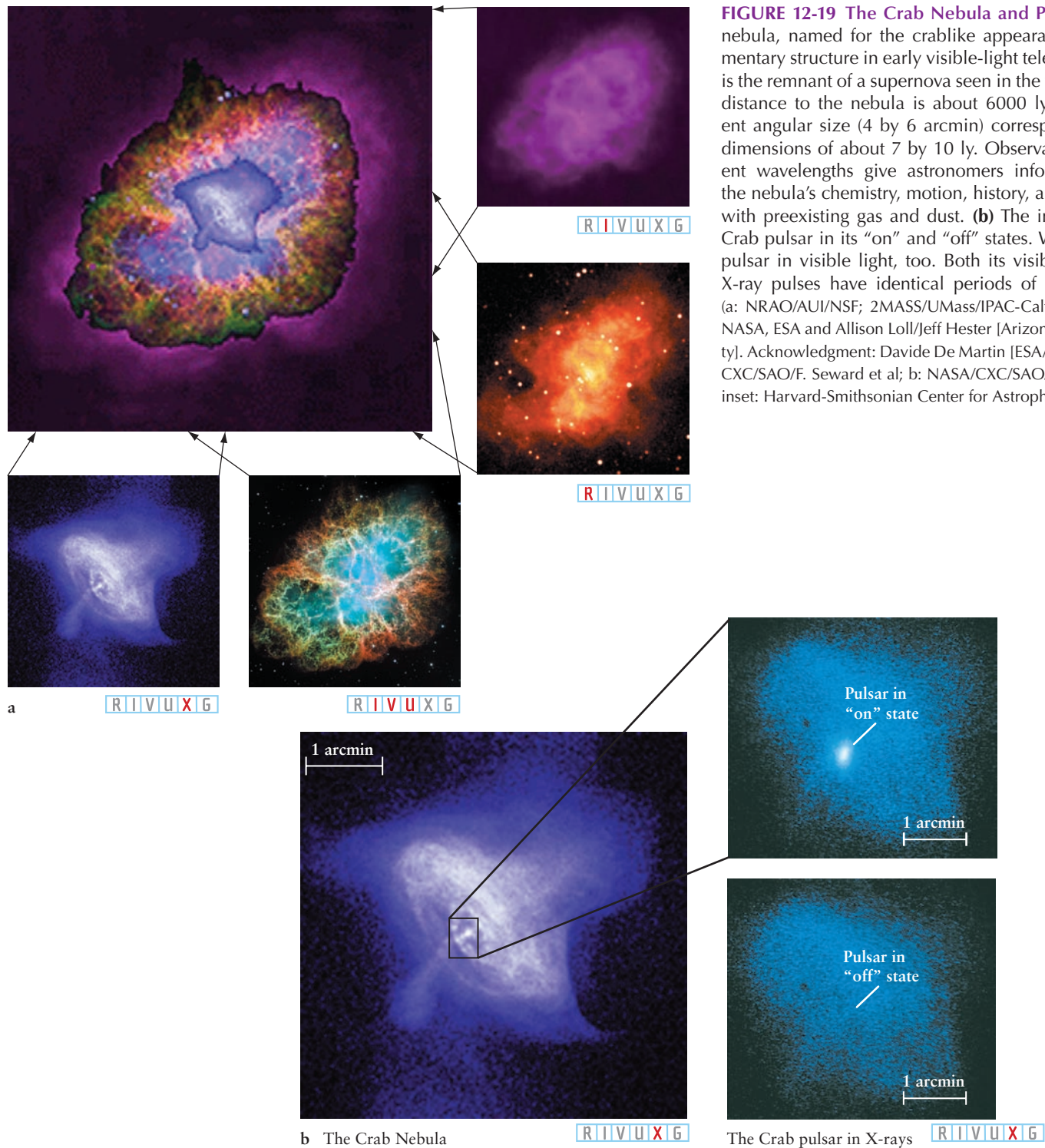


FIGURE 12-18 A Recording of a Pulsar This chart recording shows the intensity of radio emissions from one of the first pulsars to be discovered, PSR 0329+54. Note that some of the pulses are weak and others are strong. Nevertheless, the spacing between pulses is so regular (0.714 second) that it is more precise than most clocks on Earth.

Astronomers therefore looked for some recurring event in a star's life. We know that a star sheds matter as it swells to a giant. Perhaps some stars alternately expand and contract, emitting energy as they pulsate. However, this explanation also fails because any star expanding and contracting as fast as a pulsar emits pulses would explode.

Late in 1968, all controversy was laid to rest with the discovery of a pulsar in the middle of the Crab Nebula. In the year 1054, Native American, Chinese, and possibly other peoples saw and recorded the appearance of a “guest star” in the constellation of Taurus. When we turn a telescope toward this location, we find the Crab Nebula, shown in **Figure 12-19**. It looks



Focus Question 12-8

How big is a neutron star?

like the residue of an explosion and is, in fact, a supernova remnant. At the center of the Crab Nebula is the Crab pulsar (Figure 12-19b).

With expansion and contraction ruled out, the likely scenario is that pulsars are spinning objects. If the Crab pulsar is spinning, it is spinning too fast to be a white dwarf. Its period is 0.033 second, which means that it rotates 30 times each second. At that speed, something as wide as a white dwarf would immediately fly apart. Strengthening the case against pulsars being white dwarfs, astronomers discovered numerous other pulsars rotating even faster, including PSR 1937+21 (PSR for pulsar and 1937+21 for the object's approximate right ascension and declination), a stellar remnant that pulses 640 times each second. Staying with the belief that pulsars are spinning, astronomers concluded that they must be incredibly compact. Calculations revealed that, if neutron stars exist, they have a sufficiently small diameter to remain intact while rotating as fast as pulsars pulsate. Indeed, this was the *only* explanation that withstood the process of scientific scrutiny, and, as a result, the theory of neutron stars described earlier was eventually developed.

12-7 A rotating magnetic field explains the pulses from a neutron star

Why are neutron stars rotating so fast, and how do they emit pulses of electromagnetic energy? Based on observations of the rotation of the Sun, Betelgeuse, and other stars, astronomers theorize that many, perhaps most, stars have some angular momentum (rotation) as a result of their forming from swirling ensembles of gas and dust. The Sun, for example, takes nearly a full month to rotate once around its axis. While a star that explodes and forms a neutron star may have been rotating just as slowly, the rotation rate of its collapsing

neutron core increases dramatically, just as pirouetting ice skaters speed up when they pull in their arms, as shown in Figure 2-12. Recall from Section 2-7 that this occurs because the total amount of angular momentum in an isolated system (in this case, the neutron core of the star that explodes as a supernova) always remains constant. To conserve angular momentum, the core of a high-mass main-sequence star rotating once a month spins faster than once a second when it collapses to the size of a neutron star.

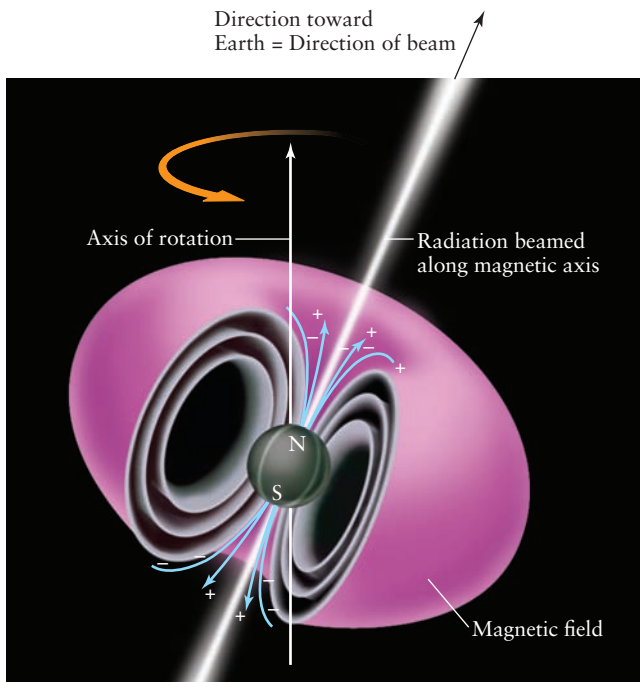
In addition to rapid rotation, most neutron stars are believed to have intense magnetic fields. In an average star like our Sun, the magnetic field is spread out over millions upon millions of square kilometers just under the star's surface (see Section 9-5). However, when the core of a star collapses down to form a neutron star, its magnetic field becomes very concentrated, in the same way as stalks of wheat in a field become concentrated when they are gathered in a bundle (Figure 12-20). The strength of a collapsing star's magnetic field increases to as much as 10^{15} times Earth's magnetic field. A magnet that strong located halfway to the Moon would lift iron rods off Earth.

5 The axis of rotation of a typical neutron star is not the same as the axis connecting its north and south magnetic poles (Figure 12-21), just as the magnetic and rotation axes of the Sun and planets (except Saturn) are different from each other. As the neutron star rotates, its powerful magnetic field therefore rapidly changes direction. Like a giant electric generator, the star creates intense electric fields, which act on protons and electrons near its surface. The powerful electric fields channel these charged particles, causing them to flow out from the neutron star's polar regions, as sketched in Figure 12-21. As the particles stream along the field, they accelerate and emit energy. The result is two very thin beams of radiation that pour out of the neutron star's north and south magnetic polar regions and sweep through space—a pulsar.

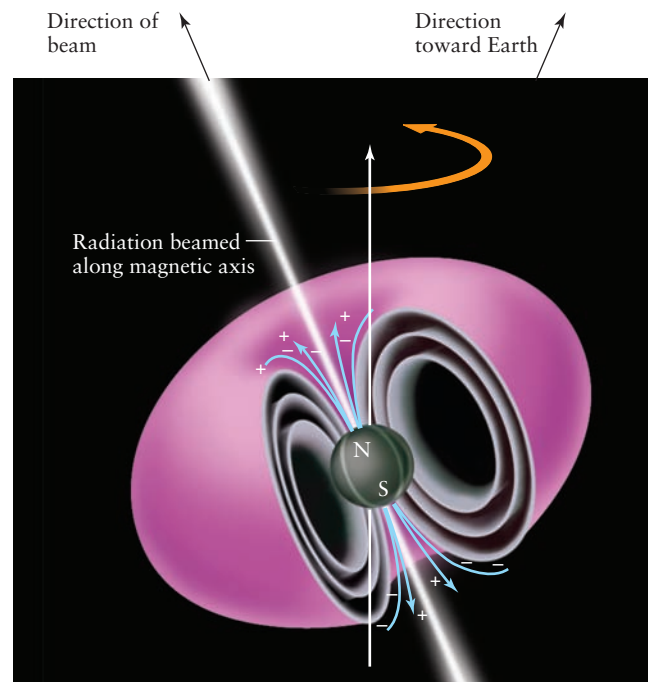


FIGURE 12-20 Analogy for How Magnetic Field Strengths Increase When growing, these wheat stalks cover a much larger area than when they are harvested and bound together.

A star's magnetic field behaves similarly. The collapsing star carries the field inward, thereby increasing its strength. (a: Sergio Pitamitz/Corbis; b: Oscar Burriel/Science Source.)



a One of the beams from the rotating neutron star is aimed toward Earth: We detect a pulse of radiation.



b Half a rotation later, neither beam is aimed toward Earth: We detect that the radiation is “off.”

FIGURE 12-21 A Rotating, Magnetized Neutron Star Calculations reveal that many neutron stars rotate rapidly and possess powerful magnetic fields. Charged particles are accelerated near a neutron star’s magnetic poles and produce

two oppositely directed beams of radiation. As the star rotates (going from **a** to **b** is half a rotation period), the beams sweep around the sky. If Earth happens to lie in the path of a beam (**a** but not **b**), we see the neutron star as a pulsar.

This explanation for pulsars is often called the **lighthouse model**. A rotating, magnetized neutron star is somewhat like a lighthouse beacon. The magnetic field causes nearby gases to accelerate and radiate. Depending on the conditions, this radiation can range from radio waves to gamma rays, and many pulsars emit in more than one spectral range. Calculations indicate that this radiation is beamed and, as the star rotates, it sweeps around the sky. If Earth happens to be located in the right direction, a brief flash can be observed each time a beam whips past our line of sight. As a result of its neutron star’s rotation, the Crab Nebula flashes on and off 30 times each second (see Figure 12-19).

The Crab pulsar is one of the youngest known pulsars, its creation having been observed some 960 years ago. Also visibly flashing is the Vela pulsar at the core of the Gum Nebula (see Figure 12-13). The Vela pulsar, with a period of 0.089 second, is the slowest pulsar ever detected at visual wavelengths. Because they use some of their energy of rotation to create the pulses they emit, *isolated pulsars slow down as they get older*. Only the very youngest pulsars are energetic enough to emit visible flashes along with their radio pulses, and the Vela pulsar was created just 11,000 years ago, which is recent in astronomical terms.

Rotating neutron stars create other phenomena besides normal pulsars Pulsars typically have magnetic fields

about 10^{12} times stronger than the magnetic field at the surface of Earth. Exceptional circumstances can conspire to raise the field strength a thousand times higher in some pulsars. If the neutron star is very hot (around 10^{11} K) and spinning very fast (100 times a second) when it forms, convection occurs throughout the neutron star, just as it does for red dwarfs (see Section 11-8). This inward and outward flow of matter, combined with the spin, generates extra magnetic fields via the dynamo effect (see Section 6-5). Combining the neutron star’s two magnetic fields yields a field 10^{15} times greater than the field at Earth’s surface. The resulting object is called a **magnetar**, and its effects are dramatic.

The magnetic fields of a magnetar are so strong that they cause the neutron star’s surface to buckle and emit stupendous bursts of X-rays and gamma rays, along with photons at other wavelengths, that last for fractions of a second. Such magnetars are called *soft gamma-ray repeaters*, or SGRs. (“Soft” means “low energy.” Even so, soft gamma rays are very powerful photons.) Interacting with surrounding gases, the magnetic fields of a magnetar rapidly slow the neutron star’s rotation rate. After 10^4 years, the body spins only once every 10 seconds, and the SGR activity subsides. Twenty-one magnetars have been observed since 1998, including

Focus Question 12-9

If the lighthouse model of pulsars is correct, do we see all nearby pulsars? Why or why not?

Focus Question 12-10

How do the rotation rates of pulsars change with time?

one in 2004 that released the brightest flash ever seen from outside our solar system. Its photons scattered off the Moon and lit up Earth's atmosphere.

In 2006, radio astronomers discovered yet another type of emission from rotating neutron stars. Called *rotating radio transients*, or RRATs, these objects give off bursts of radio and other emissions that last a few thousandths of a second, turn off for minutes or hours, and then spring briefly back to life. The neutron stars involved rotate slowly compared to most pulsars, spinning once every 0.4 to about 10 seconds. The most promising explanation, at present, is that these are older pulsars or magnetars that have slowed down.

12-8 Neutron stars have internal structure

Neutron stars are not composed solely of neutrons. Astronomers have developed a theoretical model for the surface and internal structure of neutron stars (Figure 12-22). A neutron star's interior has a radius of about 10 km (6.2 mi), with a core of superconducting protons and superfluid neutrons. A *superconductor* is a material in which electricity and heat flow without the system losing energy, whereas a *superfluid* has the strange property that it flows without any friction. Both superconductors and superfluids have been created in the laboratory. Surrounding a neutron star's core is a

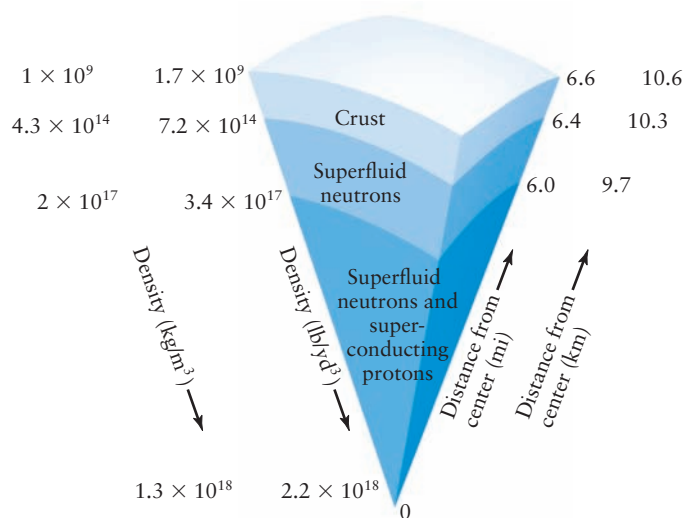


FIGURE 12-22 Model of a Neutron Star's Interior This drawing shows the theoretical model of a $1.4 M_{\odot}$ neutron star. The neutron star has a superconducting, superfluid core 9.7 km (6.0 mi) in radius, surrounded by a 0.6-km-thick (0.37-mi-thick) mantle of superfluid neutrons. The neutron star's crust is only 0.3 km (0.19 mi) thick (the length of four football fields) and is composed of heavy nuclei and free electrons. The thicknesses of the layers are not shown to scale.

layer of superfluid neutrons. The surface of the neutron star is a solid, brittle crust of dense nuclei and electrons about $\frac{1}{3}$ km thick. The gravitational force of the neutron star is so great at its surface that climbing a bump there just 1 mm high would take more energy than it takes to climb Mount Everest.

Neutron stars also have atmospheres a fraction of a meter (or of a yard) thick, as indicated by absorption lines in the spectrum of at least one of them. The absorption lines are believed to be due to scattering of radiation from the neutron star by gases in its atmosphere, just as scattering in the Sun's atmosphere causes absorption lines in its spectrum.

In 2005, astronomers found indications that some isolated neutron stars expel energy long after they form. In particular, the Cassiopeia A remnant has regions glowing from radiation deposited on them as recently as 50 years ago (some 275 years after the supernova itself). The cause of this activity is still under investigation.

As noted above, an isolated pulsar radiates energy, which causes it to slow down. This "spin down" is not always smooth. As it slows down, it becomes more spherical, and so its spinning, solid surface must readjust its shape. Because the surface is brittle, this readjustment is often sudden, like the cracking of glass, which causes the angular momentum of the pulsar to suddenly jump. Just as Earth's crust has earthquakes, calculations show that the crusts of rotating neutron stars have equivalent events. When a rotating neutron star (pulsar) undergoes such a quake, its rotation rate suddenly changes to conserve its angular momentum. This change creates a **glitch** in the rate at which the pulsar is slowing down. Such events have been observed in a number of pulsars (Figure 12-23).

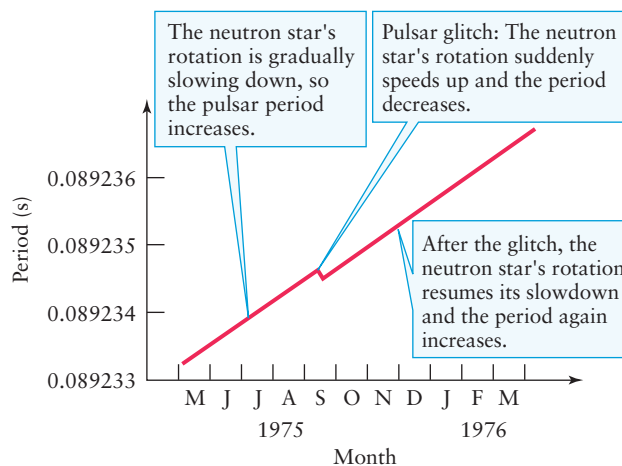


FIGURE 12-23 A Glitch Interrupts the Vela Pulsar's Spin-down Rate A glitch shown here for the Vela pulsar in 1975 changed the pulsar's rotation period.

Some pulsars are in binary and three-star systems While most pulsars appear to be isolated bodies, more than 90 have been discovered with companions (not necessarily other neutron stars). These pulsars are called *binary pulsars* and are of interest for several reasons. First, many binary pulsars pull mass onto themselves from their companions. This infall causes the pulsar to speed up (like the skater in Figure 2-12 pulling in her arms). As a result of this effect, some pulsars spin nearly a thousand times per second (716 rotations per second is the current record) and are called *millisecond pulsars*. Second, some pulsars, like the one labeled PSR 1916+13, have compact massive companions, which are often also neutron stars. Observations of binary pulsars with neutron star companions reveal that the two bodies are spiraling toward each other, as predicted by Einstein's gravitational theory, called *general relativity* (see Sections 2-1 and 12-11 for details).

In 2004, astronomers observed, for the first time, a pair of pulsars in orbit around each other (Figure 12-24). This discovery provides an exciting opportunity to further test the predictions of general relativity. For example, this theory predicts that these two pulsars should be precessing (that is, the directions that their rotation axes point should be changing; see Section 1-9) and that, like the binary pulsars just discussed, they should be spiraling in toward each other. Observations of this system in the past few years have confirmed these predictions.

In 2014, astronomers observed a millisecond pulsar, PSR J0337+1715, orbiting with two white dwarfs. This system is located in the constellation of Taurus. By

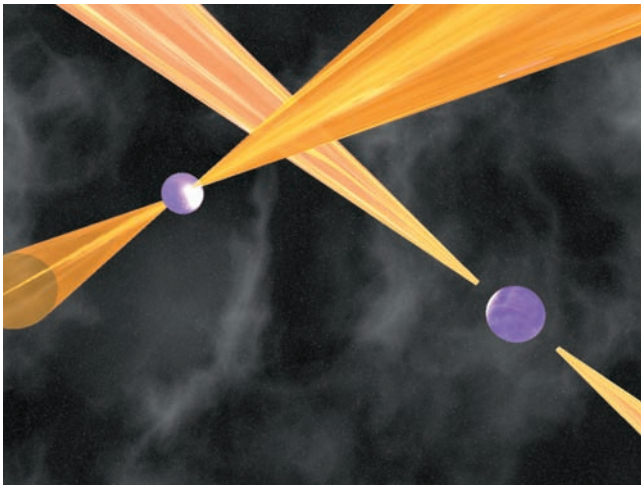


FIGURE 12-24 Double Pulsar This artist's conception shows two pulsars that orbit their center of mass. The double pulsar they represent is called PSR J0737-3039, which is about 1500 ly from Earth in the constellation Puppis. One pulsar has a 23-ms period and the other has a 2.8-s period. The two orbit once every 2.4 h. (Michael Kramer/Jordell Bank Observatory, University of Manchester)

studying it, astronomers are learning more about the gravitational interactions between closely orbiting stars and the effects of general relativity on such systems, among other things. Stay tuned!

Focus Question 12-11

Why do glitches change the rotation rates of pulsars?

12-9 Colliding neutron stars may provide some of the heavy elements in the universe

In the first few sections of this chapter we discussed nucleosynthesis, the process of creating the heavier elements. We saw how a variety of elements up to iron are created inside massive stars before the supernova occurs and are then ejected into space during the explosion. We also saw that even heavier elements are created and ejected during the supernovae themselves. However, models indicate that supernovae do not create enough of the elements heavier than iron to account for the amounts of these elements found in the universe.

Computer simulations of what happens during collisions of two neutron stars predict that, in these impacts, some of the neutrons are splashed into space, where many of them decay back into protons, fuse together, and help form various elements. The interactions that follow create a variety of heavy elements in amounts consistent with observations. For example, formation of an isotope of the element thulium requires neutron densities only available in neutron stars.

12-10 Binary neutron stars create pulsating X-ray sources

Neutron stars in binary systems may also hold the key to another regular pulse from the sky—pulses from X-ray sources. During the 1960s, astronomers obtained tantalizing X-ray views of the sky during short rocket and balloon flights that briefly lifted X-ray detectors above Earth's atmosphere. Several strong X-ray sources were discovered, and each was named after the constellation in which it was located.

Astronomers were so intrigued by these preliminary discoveries that NASA built and launched *Explorer 42*, an X-ray-detecting satellite that could make observations 24 hours a day. The satellite was launched from Kenya (to place it in an orbit above Earth's equator) in 1970, on the seventh anniversary of Kenyan independence. To commemorate the occasion, *Explorer 42* was renamed *Uhuru*, which means “freedom” in Swahili.

Uhuru gave us our first comprehensive look at the X-ray sky. As the satellite slowly rotated, its X-ray detectors swept across the heavens. Each time an X-ray source came into view, signals were transmitted to receiving stations on the ground. Before its battery and transmitter failed in early 1973, *Uhuru* had succeeded in locating 339 X-ray sources.

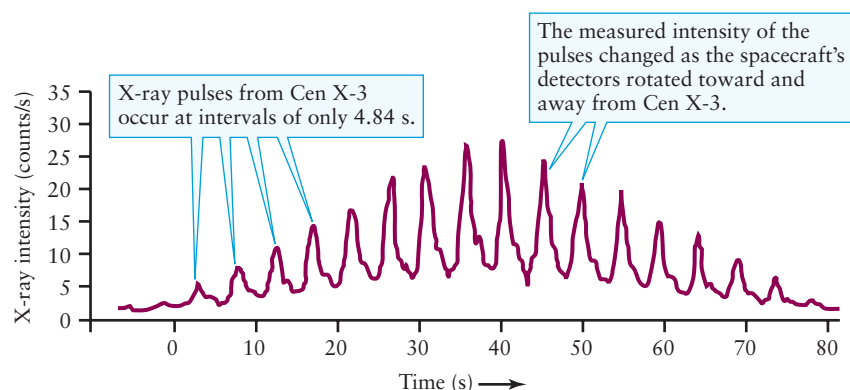


FIGURE 12-25 X-Ray Pulses from Centaurus X-3 This graph shows the intensity of X-rays detected by *Uhuru* as Centaurus X-3 moved across the satellite's field of view. The variation in the height of the pulses was a result of the changing orientation of *Uhuru*'s X-ray detectors toward the source as the satellite rotated. The short pulse period suggests that the source is a rotating neutron star.

The discovery of pulsars was still fresh in everyone's mind when the *Uhuru* team discovered X-ray pulses coming from Centaurus X-3 in early 1971. **Figure 12-25** shows data from one sweep of *Uhuru*'s detectors across Centaurus X-3. The pulses have a regular period of 4.84 seconds. A few months later, similar pulses were discovered coming from a source called Hercules X-1, which has a period of 1.24 seconds. Because the periods of these two X-ray sources are so short, astronomers suspected that they had found rapidly rotating neutron stars.

It soon became clear, however, that systems such as Centaurus X-3 and Hercules X-1 are not ordinary pulsars like the Crab or Vela pulsars. Every 2087 days, Centaurus X-3 completely turns off for almost 12 hours. This fact suggests that Centaurus X-3 is an eclipsing binary (see **Figure 10-13**), and that the X-ray source takes nearly 12 hours to pass behind its companion star.

The case for the binary nature of Hercules X-1 is even more compelling. It has an off state corresponding to a 6-hour eclipse every 1.7 days, and careful timing of the X-ray pulses shows a periodic Doppler shift every 1.7 days. This information provides direct evidence of orbital motion around a companion star. When the X-ray source is approaching us, its pulses are separated by slightly less than 1.24 seconds. When the source is receding from us, slightly more than 1.24 seconds elapse between the pulses.

Astronomers now realize that systems such as Centaurus X-3 and Hercules X-1 are examples of binary systems in which one object is a neutron star. Because all of these binaries have very short orbital periods, the distance between the ordinary star and

the neutron star must be small. This proximity enables the neutron star to capture gas escaping from the companion star.

To explain the pulsations of X-ray sources such as Centaurus X-3 or Hercules X-1, astronomers propose that the ordinary star either fills or nearly fills its Roche lobe. Either way, matter escapes from the star. If the star fills its lobe, as in the case of Hercules X-1, mass loss results from direct overflow through the Roche lobe. If the star's surface lies just inside its lobe, as with Centaurus X-3 (**Figure 12-26**), a stellar wind carries off the mass. A typical rate of mass loss for the ordinary star in such a pair is approximately $10^{-9} M_{\odot}$ per year.

The neutron star in a pulsating X-ray source, like an ordinary pulsar, rotates rapidly and has a powerful magnetic field inclined to the axis of rotation (recall **Figure 12-21**). As the gas from the companion star falls toward the neutron star, the magnetic field funnels the incoming matter down onto its magnetic polar regions. The neutron star's gravity is so strong that the gas is traveling at nearly half the speed of light by the time it crashes onto the star's surface. This violent impact creates hot spots at both poles, with temperatures of about 10^8 K, that emit abundant X-rays with a luminosity nearly 100,000 times brighter than the Sun. Observations of three such X-ray sources in 2005 revealed the impact sites on the neutron stars to range in size from 100 m to 4 km across. As the neutron star rotates, two beams of X-rays from its polar caps sweep around the sky. If Earth happens to be in the path of one of the beams, we can observe this pulsating X-ray source. The pulse period is thus equal to the neutron star's rotation period. For example, the neutron star in Hercules X-1 completes one rotation every 1.24 seconds.

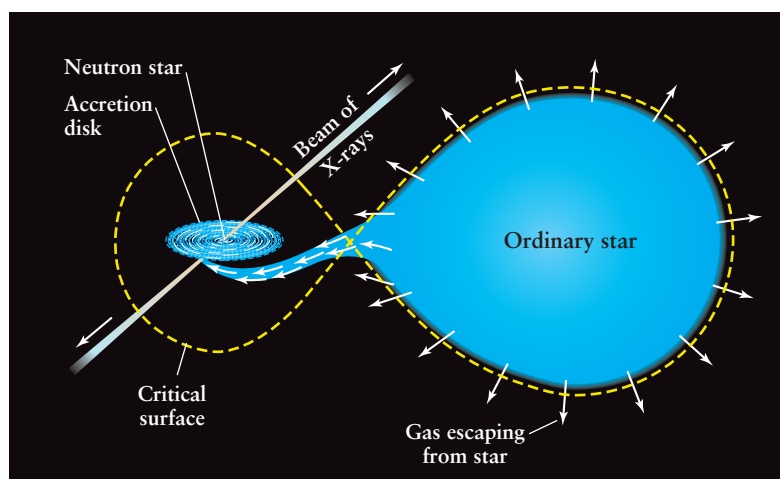


FIGURE 12-26 A Model of a Pulsating X-Ray Source Gas transfers from an ordinary star to the neutron star. The infalling gas is funneled down onto the neutron star's magnetic poles, where it strikes the star with enough energy to create two X-ray-emitting hot spots. As the neutron star spins, beams of X-rays from the hot spots sweep around the sky.

Neutron stars in binary systems can also emit powerful isolated bursts of X-rays Neutron stars can acquire additional mass from a companion star, which leads to the flaring up of the neutron stars. Beginning in late 1975, astronomers analyzing data from X-ray satellites detected sudden, powerful bursts of radiation. The record of a typical burst is shown in **Figure 12-27**. The source, called an **X-ray burster**, emits X-rays at a constant low level; then, suddenly and without warning, an abrupt increase occurs, followed by a gradual decline. A typical burst lasts for 20 seconds. Several dozen X-ray bursters have been located, most lying in the plane of the Milky Way Galaxy.

X-ray bursters, like novae, are believed to arise from mass transfer in binary star systems. With a burster, however, the stellar remnant is a neutron star rather than a white dwarf. Gases escaping from the ordinary companion star fall onto the neutron star. The energy released as this gas crashes down onto the neutron star's surface produces the low-level X-rays that are continuously emitted by the burster.

Most of the gas falling onto the neutron star is hydrogen, which becomes compressed against the hot surface of the star by the star's powerful surface gravity. In fact, temperatures and pressures in this accreting layer are so high that the arriving hydrogen is promptly converted by fusion into helium. Constant hydrogen fusion soon produces a layer of helium that covers the entire neutron star. When the helium layer is about 1 m (1 yd) thick, helium fusion ignites explosively, and we observe a sudden burst of X-rays. In other words, whereas explosive hydrogen fusion on a white dwarf produces a nova, explosive helium fusion on a neutron star produces an X-ray burster.

Just as white dwarfs have a maximum mass, above which the electron degeneracy pressure preventing their collapse gives way, so too do neutron stars have a maximum stable mass. Beyond $3.0 M_{\odot}$ the force of

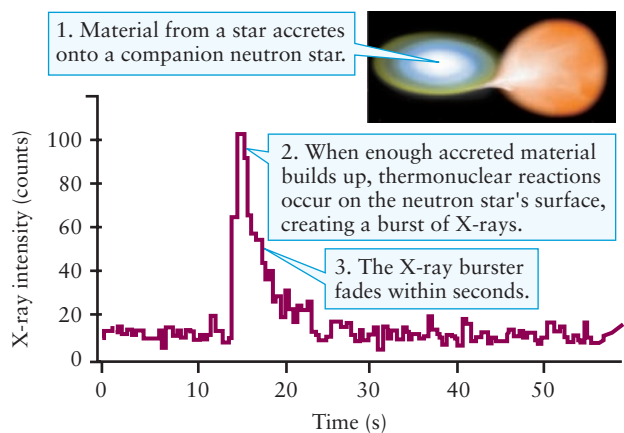


FIGURE 12-27 X-Rays from an X-Ray Burster A burster emits X-rays with a constant low intensity interspersed with occasional powerful bursts. This burst was recorded in September 1975 by an X-ray telescope that was pointed toward the globular cluster NGC 6624.

gravity overwhelms neutron degeneracy pressure and the neutron star implodes. The consequence of that event, a black hole, is what we turn to next. To understand the bizarre properties of black holes, we first must take a short excursion into the realm of Einstein's theories of relativity. Most of the concepts presented here fly in the face of our intuition. Consider, for example, a car ride.

Focus Question 12-12

Is an X-ray burster similar to a supernova in the evolution of its source?

THE RELATIVITY THEORIES

12-11 Einstein revolutionized our understanding of space, time, and gravity

It seems reasonable that if I watch you traveling at, say, 100 km/h (about 60 mi/h), my measurements of your mass, the rate at which your watch runs, and the length of your car would be the same as the measurements you make of these things. In fact, none of these measurements would be the same. At the beginning of the twentieth century, Einstein began a revolution in science by disregarding his common sense, making a few physical assumptions about nature, and using mathematics to explore their consequences. The resulting equations revealed profound, albeit highly counterintuitive, insights into how nature works. Armed with the equations that Einstein derived, scientists have been exploring and using the consequences of his theories to develop new technology and a deeper understanding of how matter, energy, space, and time interact.

Special relativity changes our conception of space and time In 1905, Albert Einstein derived his **theory of special relativity**, a description of how motion affects our measurements of distance, time, energy, and mass. He was guided by two assumptions. Although the implications of the theory proved revolutionary, the first notion seems simple:

Your description of physical reality is the same regardless of the (constant) velocity at which you move.

In other words, as long as you are moving in a straight line at a constant speed, you experience the same laws of physics as anyone else moving at any other constant speed in a straight line and in any other direction. To illustrate this first idea, suppose you were inside a closed boxcar moving smoothly in a straight line at 100 km/hr (62 mi/hr) and you dropped a pen from a height of 2 m (6.6 ft). You time how long it takes the pen to reach the floor (about 0.64 second). Then the train is stopped and you repeat the same experiment in the station. The time it would take the pen to fall would be identical.

The second idea seems more bizarre:

Regardless of your speed or direction, you always measure the speed of light to be the same.

Suppose that you are in a (very powerful) car moving toward a distant street lamp at 150,000 km/s (93,000

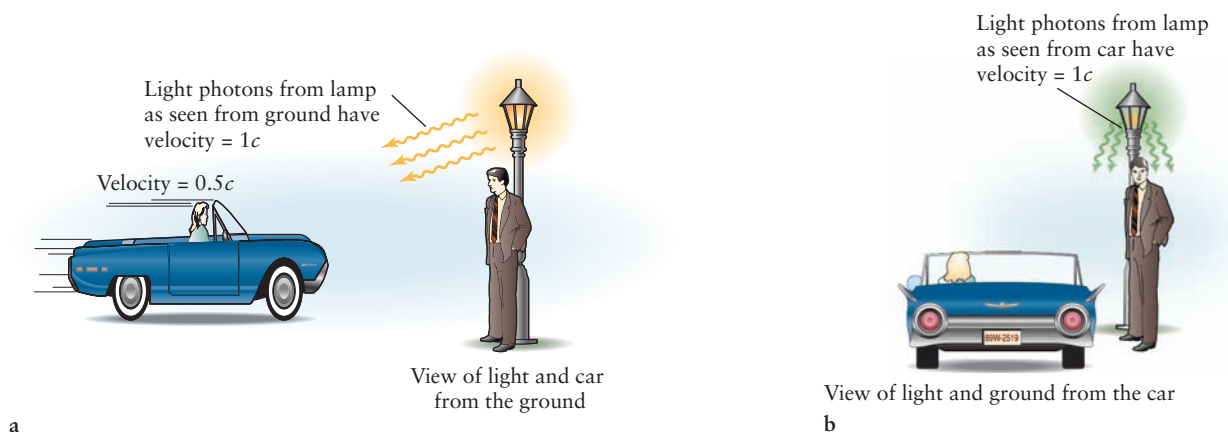


FIGURE 12-28 The Speed of Light Is Constant (a) As seen from the ground, photons of light from the lamp are traveling toward the car with a speed $v = c$ (ignoring the slight decrease in speed due to the presence of the air). (b) As seen

mi/s). In simpler terms, this is 50% of the speed of light, denoted $0.5c$ (Figure 12-28a). A friend of yours leaning on the lamppost sees you coming and also sees light from the lamp heading toward you at c . How fast do you see the photons of light coming at you? It would seem reasonable that you clock them at $c + 0.5c$, which is the speed of the photons toward you plus your speed toward them, but this is incorrect. You will, in fact, see the photons coming toward you with the same speed, c , that your friend sees them leaving the lamp (Figure 12-28b). You would, however, see that they have a different color than your friend sees. That difference is due to the Doppler shift, as explained in Section 3-2.

Einstein's theory of special relativity expresses these two assumptions mathematically, and the results of these assumptions have been confirmed in innumerable experiments. Perhaps the most well-known result of special relativity is the relationship between matter and energy, $E = mc^2$. The fact that mass is multiplied by the square of a very large number reveals that a little mass creates a lot of energy. This equation implicitly states that, contrary to most daily experience (and what most people are taught as children), matter *can* be converted into energy and energy *can* be converted into matter. This conversion takes place when stars fuse elements together, like the hydrogen fusion that occurs in the Sun today.

Three other fundamental results of special relativity also defy our everyday experience:

1. The length of an object decreases as its speed increases. In other words, if a train moves toward or away from you, you would measure its length to be shorter than you would measure its length to be when it is stopped at

the station (Figure 12-29). (This result also explains why the car in Figure 12-28a appears too short.) This result is called *length contraction*. However, if you measure the length of the moving train while you are moving with it,

from the car, moving toward the lamp at $v = 0.5c$, the photons are also traveling at the speed $v = c$. The difference in color between the light in the two figures is due to the Doppler shift.

your measurement of its length will be the same as you measured on the ground when it was at rest. The word *relativity* emphasizes the importance of the relative speed between the observer and the measured object.

2. Clocks that you see as moving run more slowly than do clocks you see at rest. This result is called *time dilation*. Indeed, the faster a clock moves relative to you, the slower it appears to tick from your perspective. For example, air travelers actually age more slowly than they would if they did not fly (although because of the relatively low speeds of aircraft travel compared to the speed of light, this difference is imperceptibly small). People flying in aircraft do not feel time passing more slowly, however, because their biological activities slow down at the same rate as the clocks around them. Only

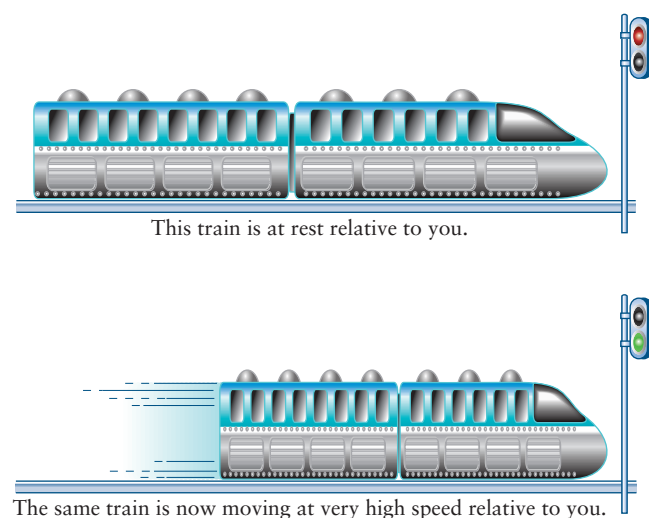


FIGURE 12-29 Movement and Space According to the theory of special relativity, the faster an object moves, the shorter it becomes in its direction of motion as observed by someone not moving with the object. It becomes infinitesimally short as its speed approaches the speed of light. The dimensions perpendicular to the object's motion are unchanged.

Focus Question 12-13

How much mass, m , would have to be destroyed to create an amount of energy, E ?



FIGURE 12-30 Curved Spacetime (a) This flat surface represents two dimensions in spacetime. In the absence of any matter in the spacetime, straight lines are straight in our intuitive sense (the sheet is flat). (b) In the presence of matter, spacetime curves, as shown by the curvature of the sheet

when mass is laid on it. Straight lines, defined by the paths that light rays take, are no longer straight in the “usual” sense. Besides changing the path of photons, this curvature also creates gravity, which (c) pulls the two masses toward each other.

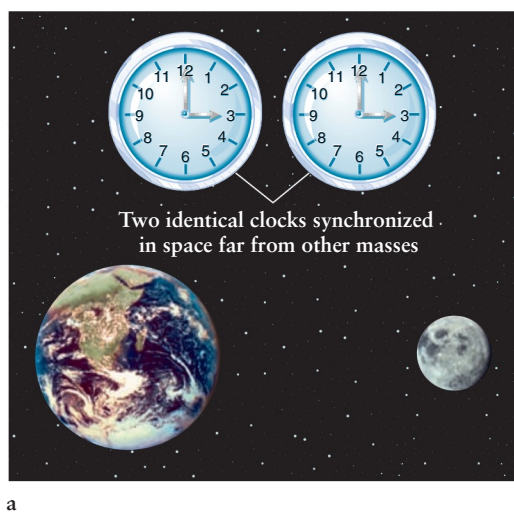
an observer moving relative to an airplane sees that the clock and the activities of the high-speed travelers aboard the plane have slowed. These connections between motion and clocks mean that space and time cannot be considered as two separate concepts. Relativity requires us to combine them in a single entity, thus creating the concept of **spacetime**.

3. The mass of an object increases as it moves faster. The concept of mass is discussed in Sections 2-5 and 2-7. The equations of special relativity reveal that an object moving at the speed of light would have an infinite amount of mass. It is impossible for any object with mass to reach this speed, much less go faster than the speed of light, because an infinite amount of mass is more than the total mass that exists in the universe. Furthermore, the equation $F = ma$ (F is the force exerted on an object, m is its mass, and a is the acceleration it undergoes; see Section 2-7) reveals that if an object’s mass is infinite, the force necessary to accelerate this object must also be infinite. Because the total force available to move matter in the universe is

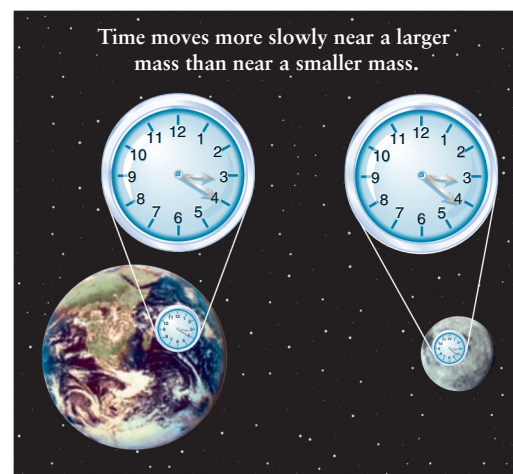
finite, no massive object can be sped up to the speed of light: The speed of light is the universal speed limit.

General relativity explains how matter warps spacetime, creating gravitational attraction While special relativity tells us how the mass of an object is related to its (constant) speed, this theory is special (in the sense of limited) because it does not account for the effects of acceleration and gravitation. These effects were incorporated by Einstein in 1915 into a more general theory, called the **theory of general relativity** (or, more simply, general relativity). It describes how spacetime changes shape in the presence of matter (Figure 12-30): the greater the mass, the more the distortion or *curvature*. Furthermore, the curvature of spacetime creates attraction between all pieces of matter in the universe (Figure 12-30c)—this attraction is what we call the *gravitational force*.

Another result of general relativity is that time slows down in the presence of matter. The greater the concentration of matter, the slower clocks tick (Figure 12-31). For example, time passes more slowly



a



b

FIGURE 12-31 Time Slows Down Near Matter (a) Two clocks in space set at exactly the same time are (b) brought to Earth and the Moon. From a vantage point far from Earth and the Moon, the clock on Earth is ticking more slowly than the

clock on the Moon. This occurs because mass slows down the flow of time, and Earth has more mass (and a higher density, which adds to the effect) than the Moon.

for us here on Earth's surface than it would for astronauts on the Moon, which has less mass than Earth. Our technology has become so sophisticated that the slowing of time due to both relative speed, as discussed earlier in this chapter, and slowing due to the presence of nearby matter, as shown in Figure 12-31, can be demonstrated in laboratories on Earth.

General relativity predicts more accurately the motion and other behaviors of matter than do Newton's laws of motion and his law of gravitation. Newton's laws are accurate only for objects with relatively small masses, slow velocities compared to the speed of light, and low densities (such as those for objects found on Earth). Newton's laws are also limited to motion sufficiently far from large masses (such as the Sun) or high-density objects (such as neutron stars). We can use Newton's laws to accurately predict the paths of projectiles on Earth and the motion of our planet around the Sun, but they do not describe the orbit of Mercury around the Sun very well (because that planet is so close to our star) or two neutron stars that orbit each other. In such cases, general relativity correctly predicts how objects move and how time passes. Closer to home, the accurate determination of location using GPS satellites orbiting Earth requires the use of the equations of general relativity.

Spacetime affects the behavior of light Besides affecting the behavior of objects, the curvature of spacetime changes the path and wavelength of

light that passes near any matter. (These behaviors are not predicted by Newton's laws.) Light travels along trajectories in space called *geodesics*. You can get an idea of how geodesics work by imagining that you are flying on the shortest possible route from

one city to another. Your path would be analogous to a geodesic for a photon, and like that particle in the presence of matter in space, you would actually be following the curve of Earth's surface, rather than going in a "straight line." The first experimental verification of general relativity, made in 1919 during a solar eclipse, was its prediction that geodesics are sufficiently curved by the mass of the Sun to cause light from stars passing nearby to arc around it (Figure 12-32).

Photons that leave the vicinity of a star also lose energy in climbing out of the star's gravitational field. They do not slow down like a bullet fired upward. Rather, they shift to longer wavelengths (Figure 12-33), an effect we see in the spectra of some white dwarfs, whose light appears redder than it would if this effect did not occur. This shift of wavelengths leaving the vicinity of a massive object is called **gravitational redshift**.

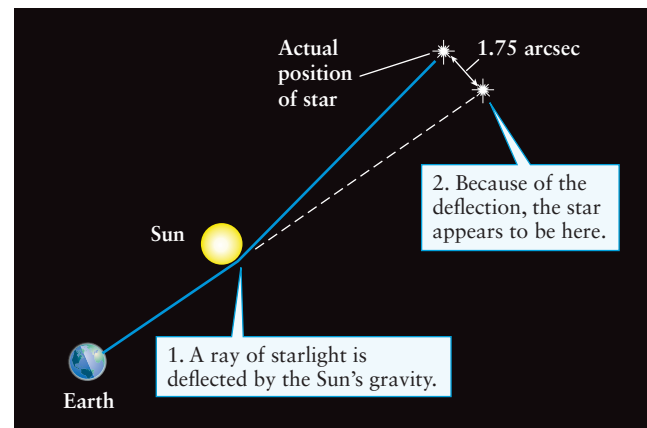


FIGURE 12-32 Curved Spacetime and the Path of Light The warping of space by matter causes light to be deflected. This was the first prediction of general relativity to be confirmed in 1919. This confirmation came when stars behind the Sun were observed during an eclipse. The star in this drawing was not observed where it was supposed to be, as a result of the Sun's gravity changing the path of its light.

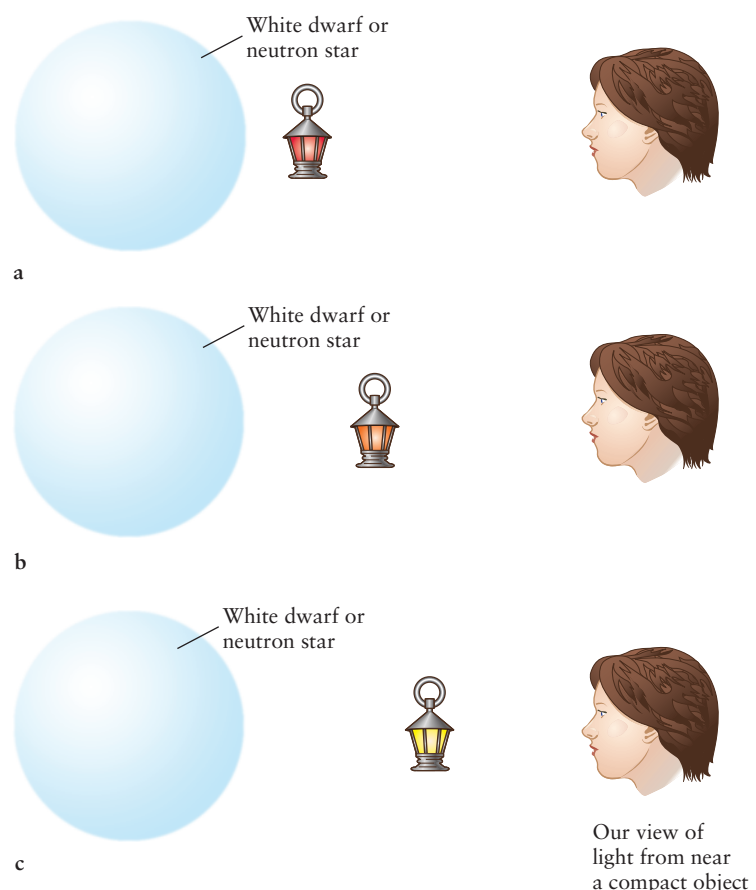


FIGURE 12-33 Gravitational Redshift The color of light from the same object located at different distances from a mass appears different as seen from far away. The photons that leave the vicinity of the massive object lose energy and are therefore redshifted. The closer the light source is to the mass, the redder the light appears and, hence, the name *gravitational redshift*. The same argument applies to light leaving the surfaces of different stars.

Focus Question 12-14

What property of matter does general relativity address that is not included in special relativity?

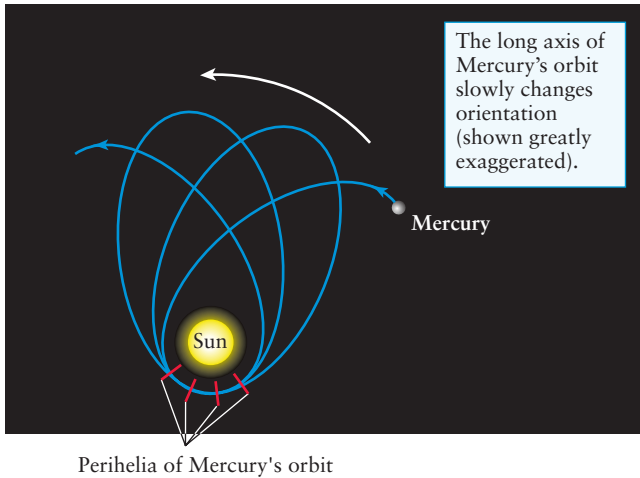


FIGURE 12-34 Mercury's Orbit Explained by General Relativity The location of Mercury's perihelion (its position closest to the Sun) and its long axis of orbit change, or precess, with each orbit. This change occurs because of the gravitational influences of the other planets plus the curvature of space as predicted by Einstein's theory of general relativity. The amount of precession is inconsistent with the prediction of the orbit made by Newton's law of gravity alone.

General relativity has been confirmed again and again, as seen in these observations:

- Light is measurably deflected by the gravitational curving of space due to the presence of matter like stars (see Figure 12-32) or entire galaxies containing billions of stars.
- The perihelion position of Mercury as seen from the Sun shifts, or precesses, by 43 arcsec per year more than is predicted by Newtonian gravitational theory (Figure 12-34). Its actual precession is exactly the amount predicted by general relativity.
- The orbits of stars in binary systems follow the paths predicted by general relativity rather than those predicted by Newtonian gravitation.
- The spectra of stars are observed to have the gravitational redshifts predicted by general relativity (see Figure 12-33).

General relativity predicts the fate of massive star cores—black holes Let us now return to stellar evolution and consider what happens when high-mass main-sequence stars (with more than $25 M_{\odot}$) explode as supernovae. As with the stars that contain between $8 M_{\odot}$ and $25 M_{\odot}$, high-mass stars become neutron stars. The neutron-star remnants of these high-mass stars, however, have more than $3 M_{\odot}$. This is problematic because above this mass, the gravitational force of the neutron star overcomes the neutron degeneracy pressure created by the interactions among its particles. As a result, such remnant stars collapse further, and no force in nature can stop their infall.

We can use the equations of general relativity to understand the fate of collapsing neutron stars. We just saw that all matter warps the space around itself (see Figures 12-30 and 12-32). When matter gets sufficiently dense, it actually causes space near it to curve so much that it closes in on itself (Figure 12-35). Photons flying outward at an angle from such a collapsing star arc back inward. Photons flying straight upward redshift so much that they lose all their energy (see Section 3-3) and hence cease to exist.

6 If light, the fastest moving of all known things, cannot escape from the vicinity of such dense matter, then nothing can! These regions out of which no matter or any form of electromagnetic radiation can leave are called **black holes**. Plummeting in on itself, a collapsing neutron star becomes so dense that it ceases to consist of neutrons. General relativity predicts that in creating a black hole, matter compresses to infinite density (and zero volume), a state called a **singularity**. However, we know that general relativity and our other theories of physics are invalid in such a situation. A more comprehensive theory of nature must be developed to explain the state of matter in a black hole's singularity. Efforts are under way to develop such a theory. The best-known nascent explanation is called *superstring theory*, but it

Focus Question 12-15

Does light change direction as it goes past Earth? Why or why not?

Focus Question 12-16

Why do neutron stars of more than about $3 M_{\odot}$ collapse to form black holes?

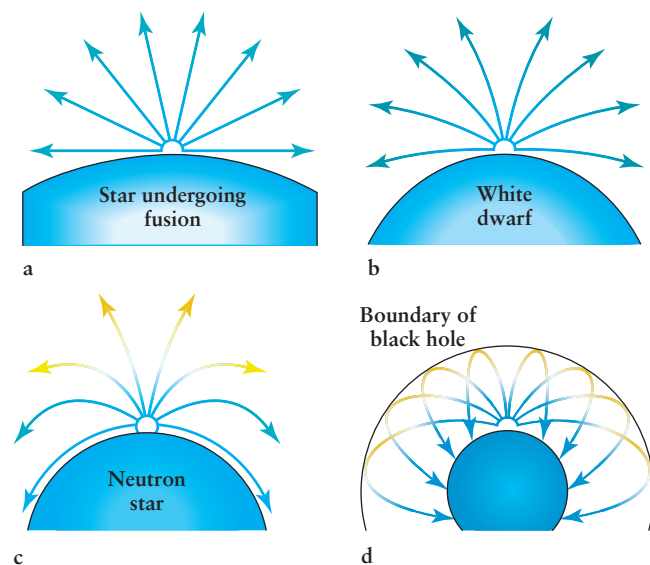


FIGURE 12-35 Trapping of Light by a Black Hole (a) The paths and colors of light rays departing from a main-sequence, giant, or supergiant star are affected very little by the star's gravitational force. (b) Light leaving the vicinity of a white dwarf curves and redshifts more, whereas (c) near a neutron star, some of the photons actually return to the star's surface. (d) Inside a black hole, all light remains trapped. Most photons curve back in, except those that fly straight upward, which become infinitely redshifted, thereby disappearing.

does not yet make testable predictions in this realm. All we can say with confidence is that the matter in black holes becomes incredibly dense and compact.

INSIDE A BLACK HOLE

The formation of a black hole is complicated, but its nature is surprisingly simple: It contains matter at its center or in a ring, it has a boundary shaped like a sphere, it either has a net electric charge or does not, and it either rotates or it does not rotate.

12-12 Matter in a black hole becomes much simpler than elsewhere in the universe

7 A black hole is separated from the rest of the universe by a boundary, called its **event horizon**. The event horizon is not like the surface of a solid or liquid body. No matter exists at this location except for the instant it takes infalling mass to cross the event horizon and enter the black hole.

We cannot look inside a black hole because no electromagnetic radiation escapes from it. Our understanding of its structure comes from the equations of general relativity. According to Einstein's theory, the event horizon is a sphere. The distance from the center of the black hole to its event horizon is called the **Schwarzschild radius** (abbreviated R_{Sch}), after the German physicist Karl Schwarzschild, who first determined its properties. The diameter of a black hole depends only on its mass. The more massive the black hole, the bigger its Schwarzschild radius is and, hence, the larger its event horizon.

According to the equations of general relativity, when a stellar remnant collapses to a black hole, it loses its internal magnetic field. The field's energy radiates away in the form of electromagnetic radiation and **gravitational radiation**. Emitted as **gravitational waves**, gravitational radiation travels as ripples in the very fabric of spacetime. Gravitational waves are also created when neutron stars or black holes collide or when stars and stellar remnants are in close orbits around each other. (Actually, gravitational radiation is emitted whenever any two things move around each other, such as a pair of dancers. When the moving bodies are less massive than stars or stellar remnants, however, we have no hope of detecting their very weak gravitational radiation with either present or projected technology.)

Gravitational radiation has not yet been directly detected, unlike, say, visible light that we “see” by its effects on our eyes or on a CCD. However, gravitational radiation has been observed indirectly by its effects on the orbits of some binary star systems. In particular, a

pair of neutron stars orbiting each other will emit so much energy as gravitational waves that the two bodies spiral toward each other. The change in their orbit is correctly predicted by general relativity. This agreement earned a Nobel Prize in Physics in 1993 for Joseph Taylor and Russell Hulse, who discovered the first binary pulsar in 1974 and measured the changes in its neutron stars' orbits.

The ripples in spacetime created by gravitational waves from stars or stellar remnants are incredibly tiny. On Earth, each meter-wide volume of space changes by less than 10^{-25} m as the waves pass by. Astronomers around the world are building *gravitational wave detectors* to measure these small changes (Figure 12-36). In operation since 2002, they have yet to directly detect any gravity waves. Astronomers expect that the next generation of gravity wave detectors will provide a direct means to observe high-energy gravitational activity in the cosmos.

Characteristics of black holes In addition to losing its internal magnetic field, matter within a black hole loses almost all traces of its composition and origin. Indeed, it retains only three properties that it had before entering the black hole: its *mass*, its *angular momentum*, and its *electric charge*. Familiar concepts, such as proton, neutron, electron, atom, and molecule, no longer apply. In addition, because few large bodies appear to have a net charge, it is also doubtful that black holes do. We therefore predict that there are only two different types of black holes: those that rotate and those that do not.



FIGURE 12-36 Laser Interferometer Gravitational Wave Observatory (LIGO) Located in Hanford, Washington, this is one of several gravitational wave (colloquially, gravity wave) detectors around the world. It has two perpendicular arms, each 4 km (2.5 mi) long. Gravity waves that pass the detector cause unequal changes in the lengths of the arms. These changes are detected by lasers inside each arm. (LIGO Laboratory)

Types of black holes If the mass creating a black hole is not rotating, the black hole that is formed does not rotate either. We call nonrotating black holes **Schwarzschild black holes** (Figure 12-37). General relativity predicts that all the mass in such a black hole collapses to a point of infinite density at its center, the singularity mentioned earlier. The rest of the volume from the event horizon to the singularity of a Schwarzschild black hole is empty space.

When the matter that creates a black hole possesses angular momentum, that matter collapses to a ring-shaped singularity located inside the black hole between its center and the event horizon (Figure 12-38). Such rotating black holes are called **Kerr black holes**, in honor of the New Zealand mathematician Roy Kerr, who first calculated their structure in 1963. Once again, the black hole is empty except for the singularity. Most Kerr black holes should be spinning thousands of times every second, even faster than the pulsars that we studied earlier in this chapter. The black hole with the highest known rotation rate, discovered in 2006, rotates a thousand times per second, consistent with this theory.

Kerr black holes cause nearby space to spiral around them like liquid in a blender (Figure 12-39). Whereas an object released from rest in the vicinity of a Schwarzschild black hole would fall straight into it, the same object released near a Kerr black hole will spiral inward. This phenomenon is relevant in our study of quasars in Chapter 13. The equations of general relativity reveal that

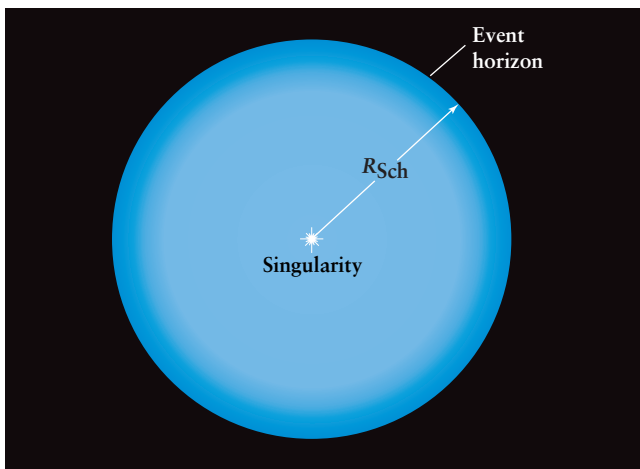


FIGURE 12-37 Structure of a Schwarzschild (Nonrotating) Black Hole A nonrotating black hole has only two notable features: its singularity and its boundary. Its mass, called a singularity because it is so dense, collects at its center. The spherical boundary between the black hole and the outside universe is called the event horizon. The distance from the center to the event horizon is the Schwarzschild radius, R_{Sch} . There is no solid, liquid, or gas surface at the event horizon. In fact, except for its location at the boundary of the black hole, an event horizon lacks any features at all.

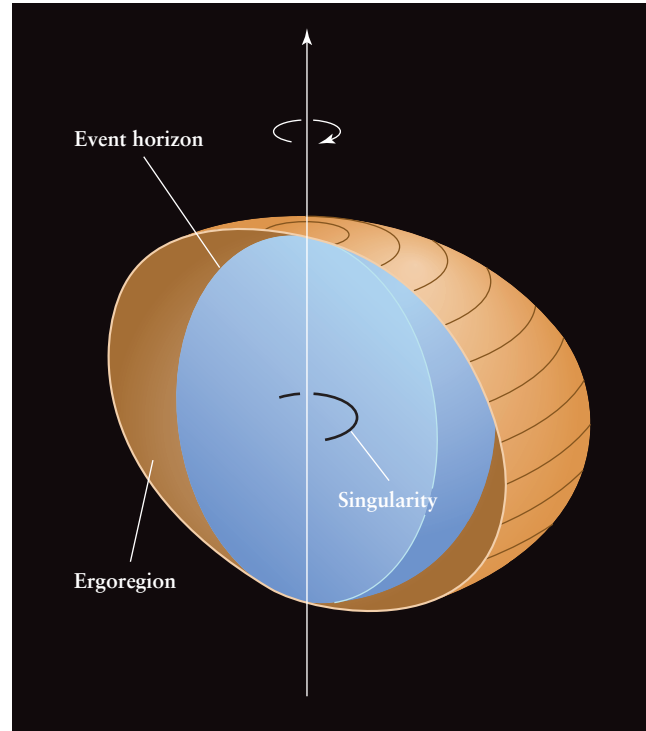


FIGURE 12-38 Structure of a Kerr (Rotating) Black Hole Rotating black holes are only slightly more complex than nonrotating ones. The singularity of a Kerr black hole is located in an infinitely thin ring around the center of the hole. The ring appears as an arc in this cutaway drawing. The event horizon is again a spherical surface. There is also a doughnut-shaped region, called the ergoregion, just outside the event horizon, in which nothing can remain at rest. Space in the ergoregion is being curved or pulled around by the rotating black hole.



FIGURE 12-39 Swirling Space in an Ergoregion Just as the liquid in this blender is being dragged around, so, too, is space dragged by a spinning (Kerr) black hole. (Ryan McVay/The Image Bank/Getty Images)

Focus Question 12-17

Where is the singularity of a rotating black hole?

a swirling region of space called the **ergoregion** (see Figure 12-38) immediately outside the event horizon of each Kerr black hole forces all objects at rest in that region to fall into the black hole.

If it is moving fast enough sideways, an object that enters the ergoregion can again fly out of it; however, if an object stops in the ergoregion, it must fall into the black hole. In 1997, matter orbiting a black hole was observed to behave in a fashion consistent with the existence of an ergoregion surrounding the hole. In the first decade of the twenty-first century, astronomers began observing spectra of atoms orbiting in the ergoregions of particularly massive black holes. The spectra showed the effects of gravitational redshift and Doppler shift predicted by Einstein's general relativity for matter orbiting in these regions.

12-13 Falling into a black hole is an infinite voyage

8 Imagine being in a spacecraft orbiting only 1000 Schwarzschild radii (15,000 km or 9300 mi) from an isolated $5\text{-}M_{\odot}$ black hole. You are held in orbit by the black hole's gravitational force. Even at that short distance, the only effect the black hole has on you is its gravitational attraction. Only when you get very, very close to the event horizon do bizarre things begin to happen. To investigate these changes, you send a cube-shaped probe with a clock toward a Schwarzschild black hole, with the same side of the cube always facing “downward” toward the black hole. The probe emits a blue glow so that you can follow its progress. What happens to the cube as it approaches the black hole?

Focus Question 12-18

How would a $1\text{-}M_{\odot}$ black hole 1 AU from Earth affect our planet?

9 From the time you launch the probe until it reaches about 100 Schwarzschild radii (1500 km or 930 mi), you see the probe descend as if it were falling toward a planet or moon (Figure 12-40a). The passage of time indicated by the probe's clock during this early stage of its infall slows down only slightly compared to your clock time farther away from the hole. At 100 Schwarzschild radii, however, the probe begins to respond significantly to the powerful tidal effect from the black hole. The face of the probe closest to the event horizon receives perceptibly more gravitational pull from the hole than its farther-away parts, and it begins to stretch apart.

By the time that the probe comes within a few Schwarzschild radii of the event horizon, the tidal forces on it are so great that it violently elongates. The part of the probe closest to the black hole accelerates downward and away from the rest of the probe. Furthermore, the sides of the probe are drawn together: They are falling in straight lines toward a common center. The net gravitational effect of moving close to the event horizon is for the probe to be pulled long and thin (see Figure 12-40). From a practical perspective, this means that the probe would be violently torn apart, because it is not composed of perfectly elastic material.

As the probe nears the black hole, the blue photons leaving it must give up more and more energy to escape the increasing gravitational force. However, unlike a projectile fired upward, photons cannot slow down. Rather, they lose energy by increasing their wavelengths. This is another example of the gravitational redshift predicted by general relativity. The closer the probe gets to the event horizon, the more its light is redshifted—first to green, then yellow, then orange, then red, then infrared, and, finally, radio waves (see Figure 12-40b, c, d).

Clock on a spaceship far from the black hole

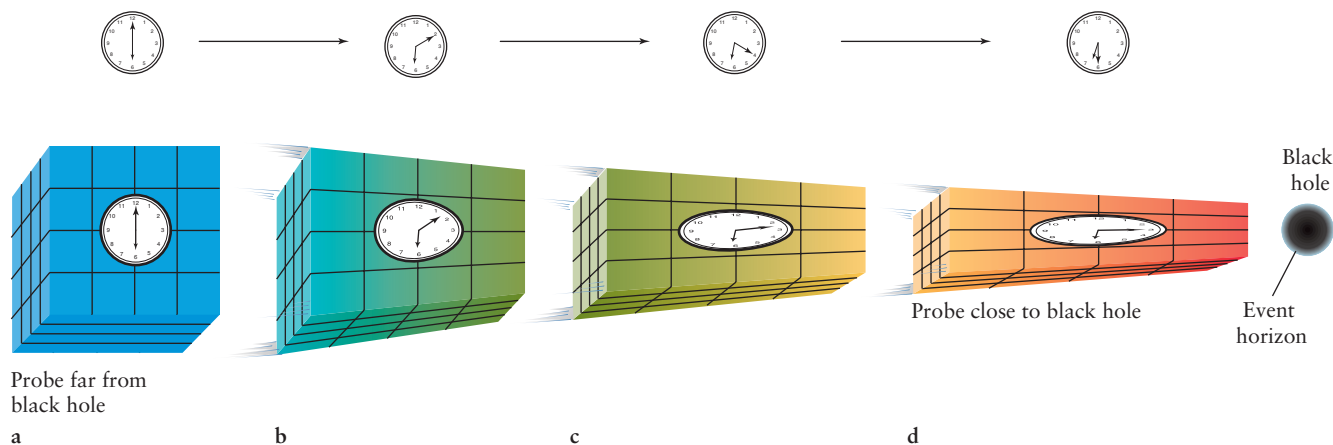


FIGURE 12-40 Effect of a Black Hole's Tidal Force on Infalling Matter (a) A cube-shaped probe 1500 km (930 mi) from a $5\text{-}M_{\odot}$ black hole. (b, c, d) Near the Schwarzschild radius, the probe is pulled long and thin by the difference in

the gravitational forces felt by its different sides. The probe changes color as its photons undergo extreme gravitational redshift and time slows down on the probe, as seen from far away.

Stranger still is the black hole's effect on time. General relativity predicts that when the probe approaches within a few Schwarzschild radii of the black hole, its infall rate will slow down as seen from far away. Also, signals from the probe show you that its clocks are running much more slowly than they did when it left your spacecraft. Time dilation becomes so great near the event horizon that the probe will appear to hover above it and its clocks will stop.

Anyone in the probe would observe something else altogether: They see the probe actually cross the event horizon and continue falling toward the black hole's singularity in a normal period of time, according to their own watch. Pulled apart by tidal effects, the probe disintegrates as it falls inward. Contrary to the science fiction concept of traveling great distances quickly by passing through a black hole, calculations indicate that objects entering them could not survive passage through, even if there were a way to come out somewhere else.

10 Could a black hole be connected to another part of spacetime or even some other universe? General relativity predicts such connections, called **wormholes**, for Kerr black holes, but astrophysicists are skeptical that the equations are correct in this regard. Their conviction is called **cosmic censorship**: Nothing can leave a local region of space that contains a singularity (that is, a black hole).

EVIDENCE FOR BLACK HOLES

12-14 Several binary star systems contain black holes

Black holes are more than fine points of relativity theory: They are real. Their presence has been observed by their effects on the orbits of other stars and on gas and dust near them, and more of them are being located all the time. To find evidence for black holes, we look first to binary star systems.

The technique for detecting black holes formed from collapsing stars is based on the interaction between the black hole and its binary companion. When one star in a close binary becomes a black hole, its gravitational attraction pulls off some of its companion's atmosphere. However, such black holes have diameters of only a few kilometers, so there is not enough room for all that gas to fall straight in. Rather, the infalling gas swirls into the black hole like water going down a bathtub drain (**Figure 12-41**). The gas that is waiting to fall in forms an **accretion disk**, that is, a disk of gas and dust spiraling in toward the black hole. Magnetic fields in the disk help pull the debris inward. Calculations reveal that this gas is compressed and thereby heated so much from the collisions of its particles that it gives off X-rays (**Figure 12-42**). Thus, if a visible star has a sufficiently tiny, sufficiently massive X-ray-emitting companion, we have located a black hole.

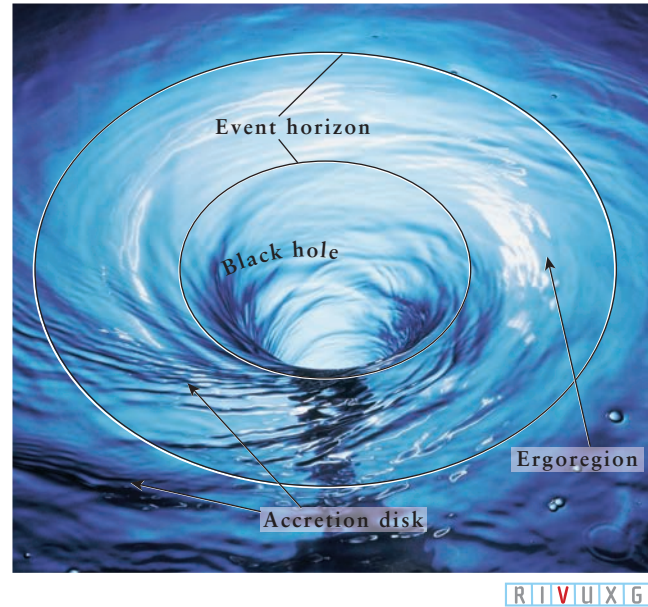


FIGURE 12-41 Formation of an Accretion Disk Just as the water in this photograph swirls around waiting to get down the drain, the matter pulled toward a black hole spirals inward. Angular momentum of the infalling gas and dust causes them to form an accretion disk around the hole. (© Chris Collins/CORBIS)

To date, in the Milky Way Galaxy alone, at least 10 stellar-remnant black holes in binary star systems have been identified (**Table 12-2**). There is also growing evidence that black holes can collide with each other or with different types of objects, such as neutron stars.

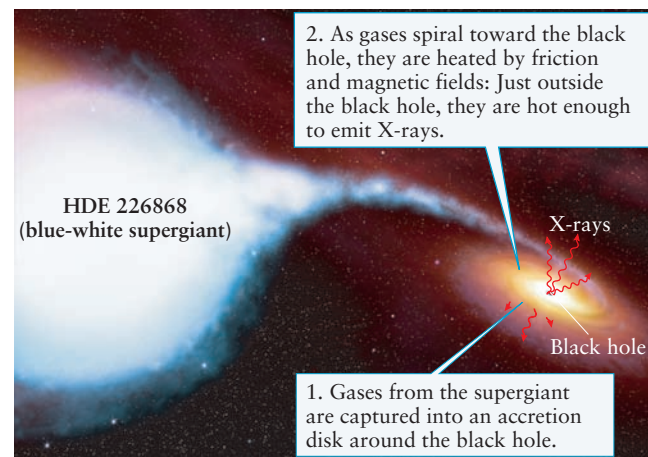


FIGURE 12-42 X-Rays Generated by Accretion of Matter Near a Black Hole Stellar-remnant black holes, such as Cygnus X-1, LMC X-3, V404 Cygni, and probably A0620-00, are detected in close binary star systems. This drawing (of the Cygnus X-1 system) shows how gas from the $30 M_{\odot}$ companion star, HDE 226868, transfers to the black hole, which has at least $11 M_{\odot}$. This process creates an accretion disk. As the gas spirals inward, friction and compression heat it so much that the gas emits X-rays, which astronomers can detect. (NASA, ESA, Martin Kornmesser [ESA/Hubble])

TABLE 12-2 STELLAR BLACK HOLES IN THE MILKY WAY

X-ray source name	Mass* of companion	Mass* of black hole
Cygnus X-1	24–42	11–21
V404 Cygni	~0.6	10–15
V461 Sgr	~10	5.5–8
GS 2000+25	~0.7	6–14
H 1705-250	0.3–0.6	6.4–6.9
GRO J1655-40	2.34	7.02
A 0620-00	0.2–0.7	5–10
GS 1124-T68	0.5–0.8	4.2–6.5
GRO J042+32	~0.3	6–14
4U 1543-47	~2.5	2.7–7.5

*Solar masses.
Data from R. Blandford and N. Gehrels, “Revisiting the Black Hole,” *Physics Today*, June 1999.

tra showed that nearby gas and stars are orbiting it extremely rapidly. They can be held in place only if the bright object contains some 3 *billion* solar masses (Figure 12-43). Given that the source’s size is only slightly larger than the solar system, it can only be a black hole.

Since 1994, black holes in the centers of many galaxies have been identified by their X-ray emissions and gravitational effects on surrounding gas and stars. For example, a distinct, Frisbee-shaped disk of gas and dust around the supermassive black hole in NGC 7052 was seen by the Hubble Space Telescope in 1998 (Figure 12-44). A black hole containing about 4 million solar masses has even been found at the center of our Milky Way, only 26,000 light-years from Earth. Indeed, evidence for central, supermassive black holes has been found in most of the several dozen nearest galaxies. To date, over 2.5 million supermassive black holes have been identified and astronomers are coming to the conclusion that they exist in the centers of most, if not all, large galaxies. We discuss the effects of these supermassive black holes (also called *active galactic nuclei*) on the galaxies in which they reside in the next chapter.

12-15 Other black holes range in mass up to billions of solar masses

Focus Question 12-19
Why is the gas in an accretion disk heated?

Based on the equations of general relativity, the fate of massive neutron stars led as early as 1939 to the idea of black holes. Since then, calculations have supported at least three other types of black holes:

1. Early in the life of the universe, black holes could have formed from the condensation of vast amounts of gas and also from the collisions of stars during the process of galaxy formation, thereby creating **supermassive black holes**, each with millions or billions of solar masses.
2. We saw in Section 11-2 that stars form in clusters. If a cluster forms with enough stars concentrated in a sufficiently small volume of space, collisions between stars in the center of the cluster may lead to the formation of a black hole with between a few hundred and a few thousand solar masses. Such objects are called **intermediate-mass black holes**.
3. Black holes could have formed during the explosive beginning of the universe as tiny amounts of matter were compressed sufficiently to form **primordial black holes**. These bodies would have masses ranging from a few grams to the mass of a planet.

Supermassive black holes In May 1994, the Hubble Space Telescope obtained compelling evidence for a black hole at the center of the galaxy M87. In the nucleus of M87 is a tiny, bright source of light. Spec-

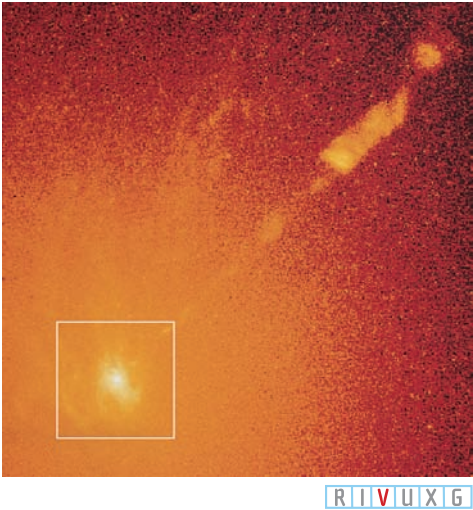


FIGURE 12-43 Supermassive Black Hole The bright region in the center of galaxy M87 has stars and gas held in tight orbits by a black hole. M87’s bright nucleus (center of the region in the white box) is only about the size of the solar system but it pulls on the nearby stars with so much force that astronomers calculate that it is a $3 \times 10^9 M_{\odot}$ black hole. One of the bright jets of gas shooting out perpendicular to the black hole’s accretion disk (discussed in Section 12-14) is visible at the upper right on this image. (Holland Ford, Space Telescope Science Institute/Johns Hopkins University; Richard Harms, Applied Research Corp.; Zlatan Tsvetanov, Arthur Davidsen, and Gerard Kriss at Johns Hopkins; Ralph Bohlin and George Hartig at Space Telescope Science Institute; Linda Dressel and Ajay K. Kochhar at Applied Research Corp. in Landover, Md.; and Bruce Margon from the University of Washington in Seattle; NASA)

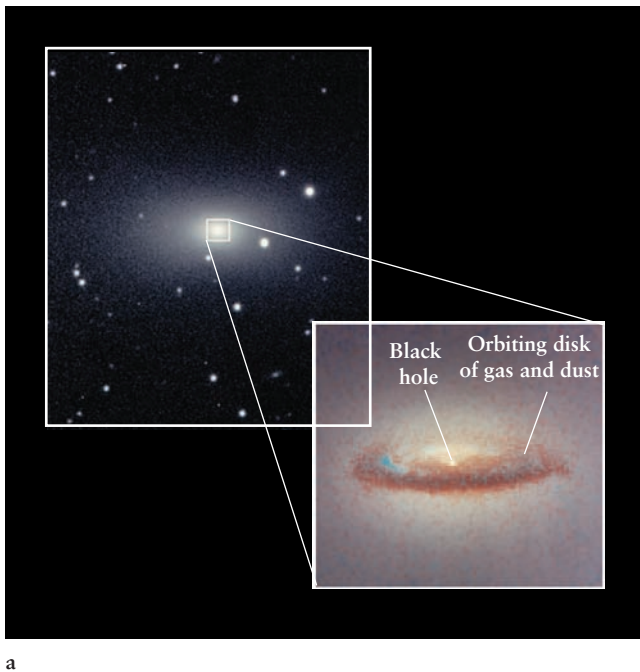
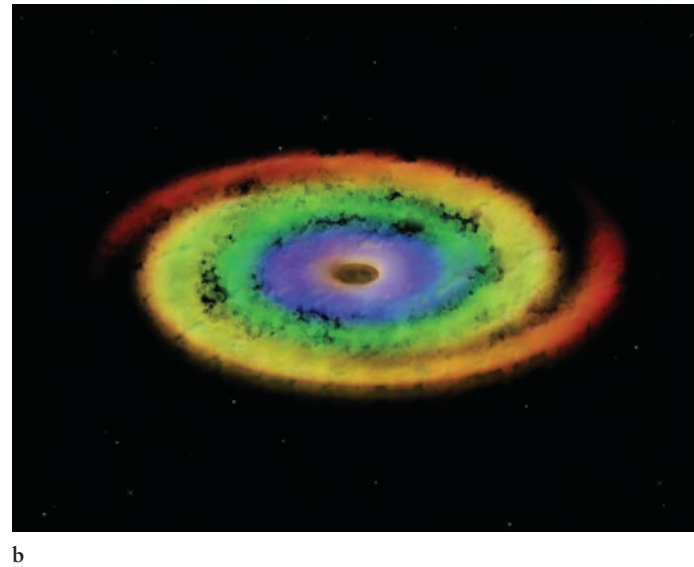


FIGURE 12-44 Disk of Gas and Dust Around a Supermassive Black Hole (a) Swirling around a $3 \times 10^8 M_{\odot}$ black hole in the center of the galaxy NGC 7052, this disk of gas and dust is 3700 ly across. The gas is cascading into the black hole, which will consume it over the next few billion years. The black hole appears bright because of light emitted by the hot, accreting gas outside its event horizon. NGC 7052



is 191 million ly from Earth in the constellation Vulpecula. (b) This drawing shows how the gases spiraling inward in an accretion disk heat up as they approach the black hole. Color coding follows Wien's law: red (coolest), followed by orange, yellow, green, blue, and violet (hottest). (a: R. P. van der Marel, STScI; F. C. van den Bosch, University of Washington; NASA; b: NASA/CXC/SAO)

Earlier in this book we studied how tidal effects of one body on another occur in several situations, including the Earth–Moon system and the heating of Io and Europa by Jupiter. In 2004, astronomers observed several supermassive black holes in other galaxies absorbing parts of passing stars. These black holes created such high tides on those stars that the stars were pulled apart. Some of each star's mass was then drawn into the black hole. These discoveries were made because infalling gas from the stars was rapidly heated to millions of kelvins, causing the gas to emit a burst of X-rays that was observed by orbiting telescopes. Subsequent observations also reveal that supermassive black holes can consume vast quantities of interstellar gas, sometimes stripping enough gas from a galaxy to prevent it from forming new generations of stars.

As we will explore in more detail in Chapter 13, galaxies were created from condensing gas and stars in the early universe. Our technology is now providing us with observations of that epoch of the universe, and the formation process of supermassive black holes in the centers of galaxies is now being studied. These black holes apparently formed, in part, from the gas that was condensing to create galaxies and, in part, from the collisions of stars and black holes that came into existence shortly after the beginning of time. Observational evidence indicates that the formation of supermassive black holes is still ongoing. We will explore this further later in this chapter.

Intermediate-mass black holes At the end of the twentieth century, astronomers began discovering objects that appear to be black holes with between 10^2 and $10^5 M_{\odot}$. Again, these objects were identified as potential black holes by the intensity and spectra of the X-rays that they emit and by the orbits of stars around them. For example, near the center of the galaxy M82, astronomers found what appears to be a $500\text{--}1000 M_{\odot}$ black hole (Figure 12-45), while in the galaxy NGC 1313, astronomers have observed what appears to be two intermediate-mass black holes, each with $200\text{--}500 M_{\odot}$. Most intermediate-mass black holes are observed in globular clusters.

In support of the belief that these objects are black holes, computer simulations of stars in the crowded regions, where such objects are found, show that black holes in this mass range can form as a result of collisions and close tidal interactions between stars. Do not fear—such a black hole will not develop near Earth. There have to be at least a million times more stars per volume than there are in our stellar neighborhood for such frequent stellar collisions or near misses to occur.

Primordial black holes Even more exotic black holes may have formed along with the universe itself. The British astrophysicist Stephen Hawking has proposed that the Big Bang explosion from which most astronomers believe the

Focus Question 12-20

Where are supermassive black holes located?

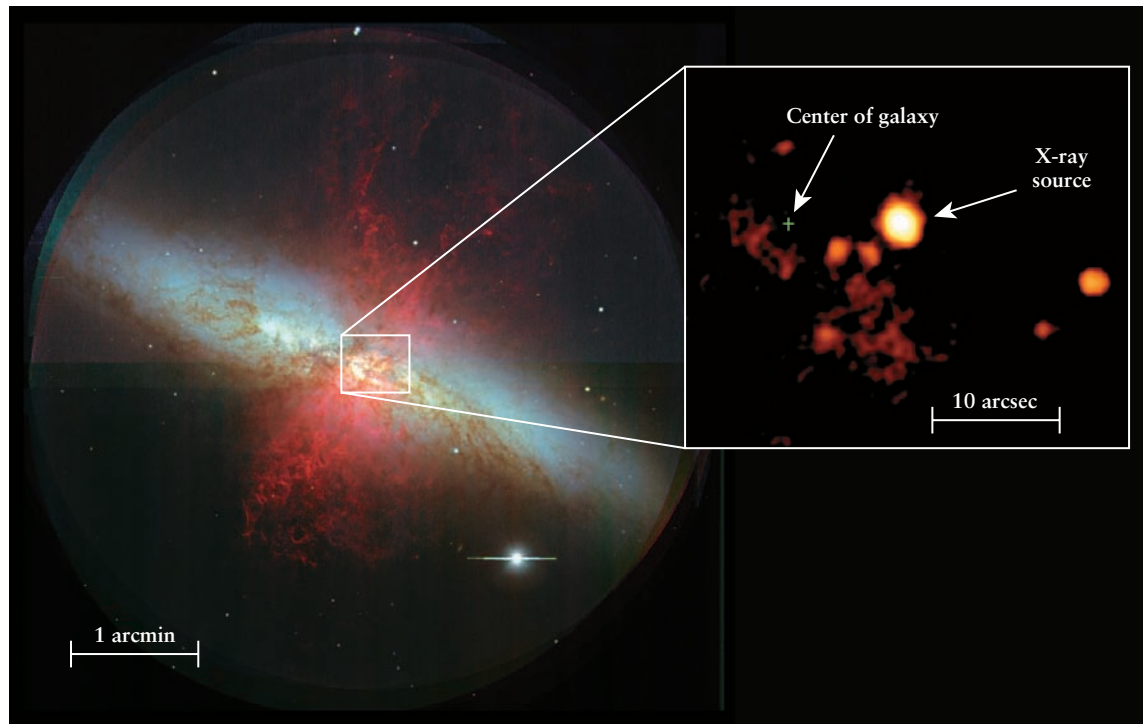


FIGURE 12-45 An Intermediate-Mass Black Hole M82 is an unusual galaxy in the constellation Ursa Major. The inset shows an image of the central region of M82 from the Chandra X-ray Observatory. The bright, compact X-ray source shown

varies in its light output over a period of months. The properties of this source suggest that it is a black hole of roughly $500 M_{\odot}$. (NASA, ESA, and the Hubble Heritage Team [STScI/AURA]; inset: [NASA/SAO/CXC])

universe emerged may have been chaotic and powerful enough to have compressed tiny knots of matter into primordial black holes. Their masses may have ranged from a few grams to greater than the mass of Earth. Astronomers have not yet observed evidence of primordial black holes, although that does not mean they do not necessarily exist—just that we cannot yet detect them.

12-16 Black holes and neutron stars in binary systems often create jets of gas

We saw in Section 12-10 how rotating beams of radiation are emitted by neutron stars. Similarly, many neutron stars and black holes emit steady (nonrotating) pairs of gas jets shooting out in opposite directions.

Consider a black hole (neutron stars behave analogously) in a binary system. If the companion star is still fusing material, its outer layers can be pulled off, as discussed in Section 11-15. The infall of this matter toward the compact companion (which has a Schwarzschild radius of only a few kilometers or miles) is too rapid for all of the mass to immediately enter the event horizon. The resulting accretion disk around the black hole is the key to explaining the presence of the jets.

As the disk mass spirals down toward the event horizon, this gas is compressed into a smaller volume. Such compression heats the gas, which, in turn, causes its pressure to increase. As a result of the increased

pressure, the gas expands, forming a doughnut-shaped region around the black hole (**Figure 12-46**). As the gas starts the final plunge toward the black hole, its temperature skyrockets to tens or even hundreds of millions of kelvins. At such temperatures, the pressure is so great that much of the infalling gas expands and,

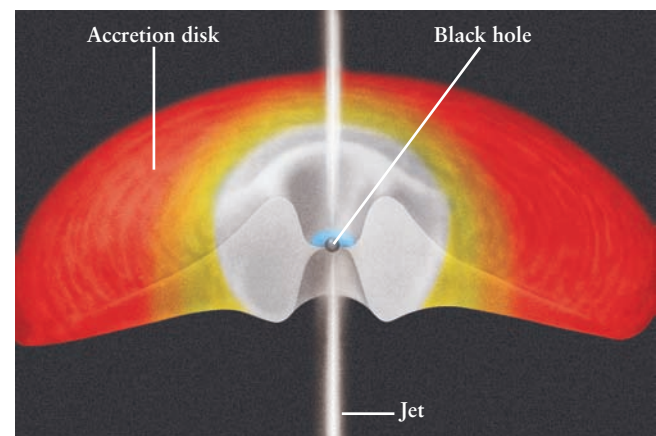


FIGURE 12-46 Jets Created by a Black Hole in a Binary System Some of the matter spiraling inward in the accretion disk around a black hole is superheated and redirected outward to produce two powerful jets of particles that travel at close to the speed of light. The companion star is off to one side of this drawing.

finding little resistance perpendicular to the plane of the accretion disk, it squirts out as two jets (Figure 12-46). These jets are prevented from spreading out by a magnetic field created as the hot gases orbit in the accretion disk and by pressure from surrounding gas. Astronomers have measured the speed of this gas to be as fast as 32×10^6 km/hr (20×10^6 mi/hr) or 3% of the speed of light.

Whereas neutron stars and stellar-mass black holes create jets as a result of being in binary star systems, supermassive black holes have so much gravitational attraction that they pull huge quantities of nearby interstellar gas and dust into orbit around themselves without needing a binary companion. Figure 12-44a shows such a disk, and Figure 12-43 shows one of the jets emitted by the supermassive black hole in M87. We will explore more about jets created by black holes in Chapter 13.

GAMMA-RAY BURSTS

Some black holes, neutron stars, and the supernovae that accompany their formation are apparently the sources of the most energetic and enigmatic energy emissions known in the universe. Called **gamma-ray bursts**, these pulses of gamma rays (the shortest-wavelength electromagnetic radiation) emit more energy than supernovae from even smaller volumes of space.

12-17 Gamma-ray bursts are the most powerful explosions in the known universe

Because atomic bomb blasts emit pulses of gamma rays, the U.S. military put the gamma-ray-detecting *Vela* satellites in orbit around Earth to monitor illegal nuclear explosions in the 1960s. In 1973, astronomers were told that since 1967, these satellites had been detecting bursts of gamma rays from objects in space. This news intrigued the astronomical community because, at that time, there was no known mechanism to emit such radiation.

Gamma-ray bursts each last between a few milliseconds and about 1000 seconds (just over 16 minutes). There appear to be at least three types of such bursts. One group typically lasts for a few tenths of a second. The second group typically lasts for about 40 seconds. The third group appears to have properties of both of the first two groups; hence, it is called a *hybrid gamma-ray burst*. Unlike X-ray bursts (see Section 12-10), each gamma-ray burst occurs only once.

In 2004, astronomers detected the first remnant of a gamma-ray burst that occurred in the Milky Way. In 2008, they observed a gamma-ray burst 7.5 billion light-years away that also emitted enough visible light to be (just barely) visible to the naked eye here on Earth. For comparison, it was as bright as the Triangulum Galaxy, which contains some 30 billion stars and is only 3 million light-years away. More than 5000 gamma-

ray bursts have been observed, and more are being discovered at a rate of about one per day. Plotting their locations on the celestial sphere (Figure 12-47), astronomers have discovered that the shorter-lived bursts occur slightly more commonly in the plane of the Milky Way Galaxy than elsewhere, while the longer-lived ones occur randomly over the sky.

Since 1997, optical and infrared “afterglows” of many gamma-ray bursts have been detected (Figure 12-48). These often persist for days, allowing astronomers time to study their spectra in detail. Such observations reveal that the light has passed through intergalactic gas clouds. Combining this discovery with observations of the bursts’ spectra and redshifts reveals that the longer-lasting bursts typically originate more than 6 billion light-years away, far outside our Galaxy. The most distant gamma-ray burst observed so far occurred 13.1 billion light-years away. Many of the shorter bursts occur in more nearby galaxies. Taking their distance into account, a typical long-lived gamma-ray burst emits as much energy in 100 seconds as the Sun will emit over the 10 billion years that it will be on the main sequence. Gamma-ray burst afterglows emit 2.5 million times as much visible light as does the brightest supernova ever observed.

Deep-sky observations of some gamma-ray bursts have begun to show galaxies in the same locations where

Focus Question 12-21

Why doesn’t all the gas in a black hole’s accretion disk enter the black hole?

Focus Question 12-22

How long do gamma-ray bursts last?

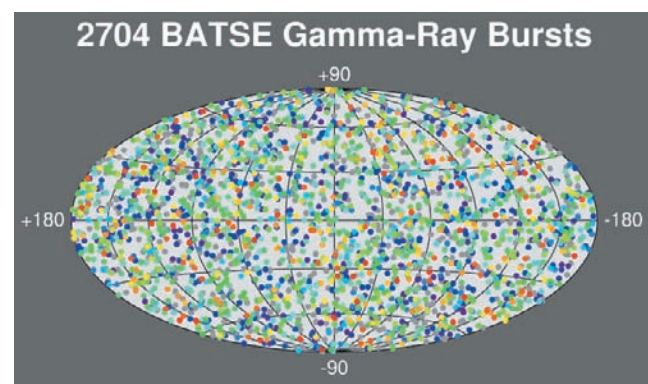


FIGURE 12-47 The Most Powerful Known Bursts Gamma-ray bursts have been observed everywhere in the sky, indicating that, unlike X-ray bursters, most do not originate in the disk of the Milky Way Galaxy. This map of the entire sky “unfolded” onto the page shows 2704 bursts detected by the Burst and Transient Source Experiment (BATSE) aboard the Compton Gamma-ray Observatory. (The colors indicate the brightness of the bursts; they are coded brightest in red, dimmest in violet. Gray dots indicate incomplete information about the burst strength.) (GSFC/NASA)

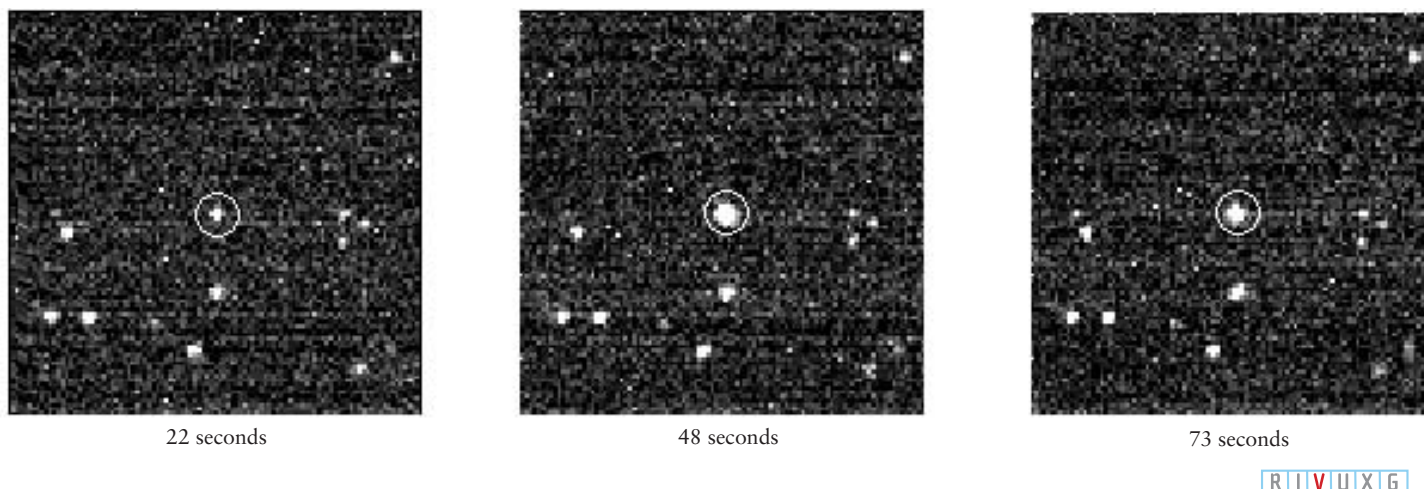


FIGURE 12-48 Gamma-Ray Bursts In 1999, astronomers photographed the visible-light counterpart of a 100-second gamma-ray burst some 9 billion light-years away in the constellation Boötes.

The times indicated are after the burst began. (Carl Akerlof, University of Michigan, Los Alamos National Laboratory, Lawrence Livermore National Laboratory)

the bursts were discovered. As **Figure 12-49** reveals, the bursts do not appear in the centers of their host galaxies. Likewise, the bursts in the plane of the Milky Way are not located near the galactic nucleus. These observations indicate that bursts are not associated with the supermassive black holes in the galactic cores. Research suggests that gamma-ray bursts are instead associated with stellar cataclysms.

In 2003, observations by NASA's *High-Energy Transient Explorer* observed a 30-second gamma-ray burst that was seen to occur at the same place as a particularly powerful supernova, called a *hypernova*, occurring in **Wolf-Rayet stars**. These are rotating stars of at least $20 M_{\odot}$ with strong magnetic fields and stellar winds. The bursts associated with these stars occur when gamma rays and a lot of gas are funneled out along the rotation axis of the star prior to its exploding. If we are looking down the axis from Earth, the gamma rays create the burst, and the gas, traveling more slowly, creates the subsequent afterglow. Observations reveal that hypernovae and their associated gamma-ray bursts are extremely messy affairs, with powerful aftershocks, rather than a single, smooth outburst. Because gamma-ray bursts associated with supernovae occur near the beginning of these events, the bursts enable astronomers to watch the supernovae unfold, providing further insights into these explosions and the mechanism by which they generate gamma-ray bursts.

Other gamma-ray bursts occur when black holes collide, when a black hole swallows a neutron star, or when a pair of neutron stars collide and form a black hole. Observations made in 2005 confirm calculations predicting that the collision of such pairs of objects each lead to a rapid, intense burst of gamma rays.

To better understand gamma-ray bursts, astronomers have constructed telescopes for use both on Earth and in orbit that are able to rapidly slew (change direction) to look in the direction in which these events occur in

space. These telescopes, including the *Swift* Gamma-Ray Burst Mission and the *High Energy Transient Explorer 2*, see these bursts and provide new insights into their origins and dynamics. *Swift* carries not only a gamma-ray telescope to locate the bursts but also separate X-ray and optical telescopes to study their afterglows.

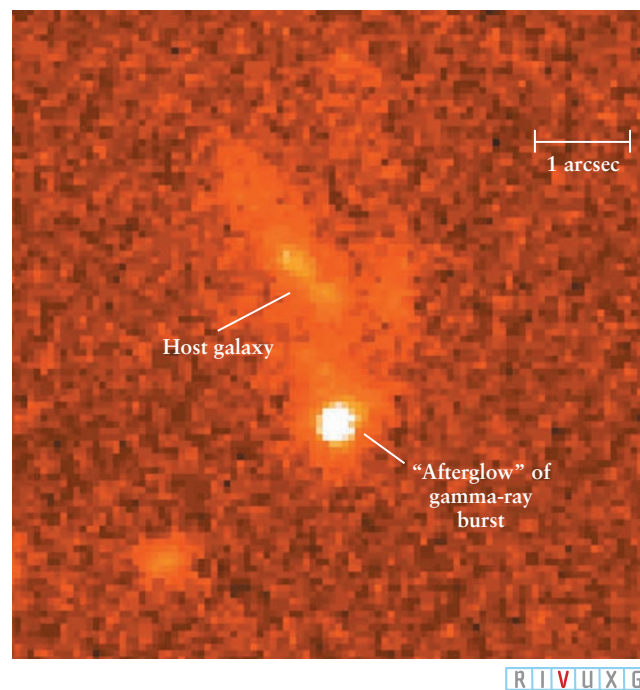


FIGURE 12-49 The Host Galaxy of a Gamma-Ray Burst The Hubble Space Telescope recorded this visible-light image in 1999, 16 days after a gamma-ray burst was observed at this location. The image (color-coded to show contrast, rather than true color) reveals a faint galaxy that is presumed to be the home of the gamma-ray burst. The galaxy actually has a bright blue color, indicating the presence of many young stars. (Andrew Fruchter, STScI; and NASA)

12-18 Black holes evaporate

In exploring the fate of black holes, astrophysicists find that, once again, these objects confound common sense—they evaporate! With a black hole cloaked behind the event horizon and its mass collapsed into a singularity, it must seem that there is no way of getting mass from the black hole back out into the universe again. No way, that is, until you recall that mass and energy are two sides of the same coin. What if there was a way of converting the black hole's mass into a form of energy that *could* get out, such as gravitational energy? The transformation does occur, according to an idea first proposed by Stephen Hawking.

Black holes convert their mass into energy by a process called **virtual particle** production. If things in this chapter aren't weird enough, consider this: The laws of nature allow pairs of particles, called virtual particles, to spontaneously appear. Each pair consists of a particle and its anti-particle—such as a proton and an antiproton, an electron and an antielectron (positron), or a pair of photons. The particles in each pair are identical, except for having opposite electric charges, and they annihilate each other and normally disappear without a trace within a very short period of time, typically 10^{-21} seconds. Calculations reveal that billions and billions of pairs of virtual particles are seething in and out of existence in your body (and every other equivalent volume of space) every second.

We know that virtual particles actually flash into existence spontaneously throughout the universe because some have been made into real particles in the laboratory. This creation is done using high-speed particles in particle accelerators that slam into the pair of virtual particles and push them apart before they annihilate each other. The mass or energy given to the virtual particles in making them real comes from the laboratory particles, which lose an equal amount of energy.

Now, suppose two virtual particles form *just outside* the event horizon of a black hole. They would experience an exceptionally powerful gravitational field. If one particle is created slightly farther from the hole than its companion, the two virtual particles feel a tidal force from the black hole's tremendous gravitational pull. If the gravitational tidal force is strong enough, it can pull the two virtual particles apart before they can annihilate each other. This separation makes them real (**Figure 12-50**).

11 To conserve momentum, at least one of the newly formed real particles always falls into the black hole. But the other particle will sometimes have enough kinetic energy to escape from the vicinity of the black hole. This latter particle flies free into space. Because the gravitational energy used to make the virtual particles real came from the mass of the black hole, the hole loses mass equal to $E = mc^2$, where m is the mass of the freed particle and c is the speed of light or, if the pair of particles are photons, $E = hc/\lambda$, where λ is the wavelength of the departing photon and h is Planck's constant. The energy from inside the black hole is transmitted outside

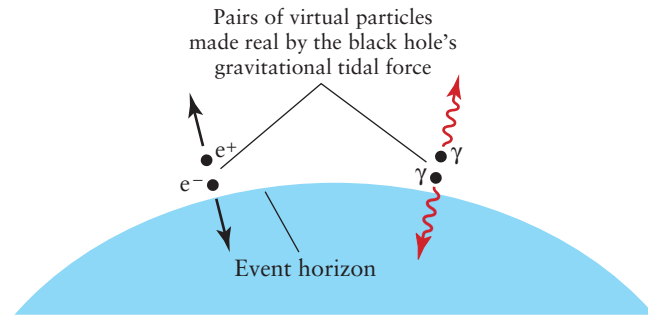


FIGURE 12-50 Evaporation of a Black Hole Throughout the universe, pairs of virtual particles spontaneously appear and disappear so quickly that they do not violate any laws of nature. The tidal force just outside the event horizon of a black hole is strong enough to tear apart two virtual particles that appear there before they destroy each other. The gravitational energy that goes into separating them makes them real and, therefore, permanent. At least one of each pair of newly created particles falls into the black hole. Sometimes the other particle escapes into the universe. Here, we see just a few particles in the making: an electron (e^-), a positron (e^+), and a pair of photons (γ).

the event horizon as gravitational radiation to replace the energy taken away by the newly formed particles. This is called the **Hawking process**, and the particles leaving the vicinity of the black hole are called *Hawking radiation* (even though many of the particles are not electromagnetic radiation). As a result of the Hawking process, the black hole loses mass—it evaporates.

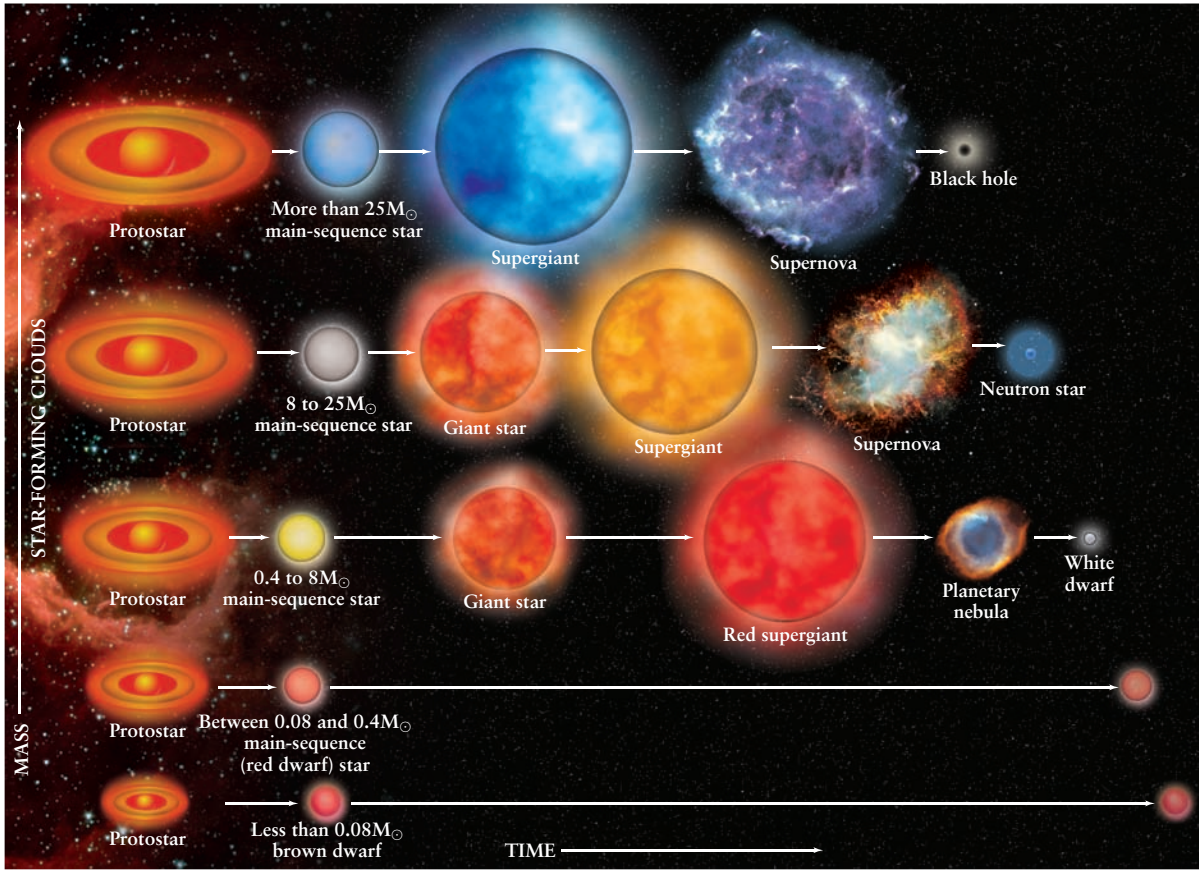
The time it takes a black hole to completely evaporate by the Hawking process increases with the mass of the hole. More massive black holes have lower tidal forces at their event horizons than do less massive and, therefore, smaller-diameter black holes. Therefore, the *rate* at which virtual particles are converted into real particles around bigger black holes is actually slower than it is around smaller ones. The faster the real particles are produced, the faster they wear down the mass of the black hole; thus, lower-mass black holes evaporate more rapidly than higher-mass black holes. Indeed, calculations indicate that in its final moments of evaporation, a black hole should create particles so quickly that it should appear to explode.

A stellar black hole of $5 M_\odot$ would take more than 10^{62} years to evaporate (much longer than the present age of the universe), whereas a 10^{10} kg primordial black hole (equivalent to the mass of Mount Everest) would take only about 15 billion years to evaporate. Since the universe is about 13.8 billion years old, a black hole of this size and age is well on its way to completely evaporating. Astronomers are trying to identify events in space that correspond to the violent, particle-generating deaths of primordial black holes.

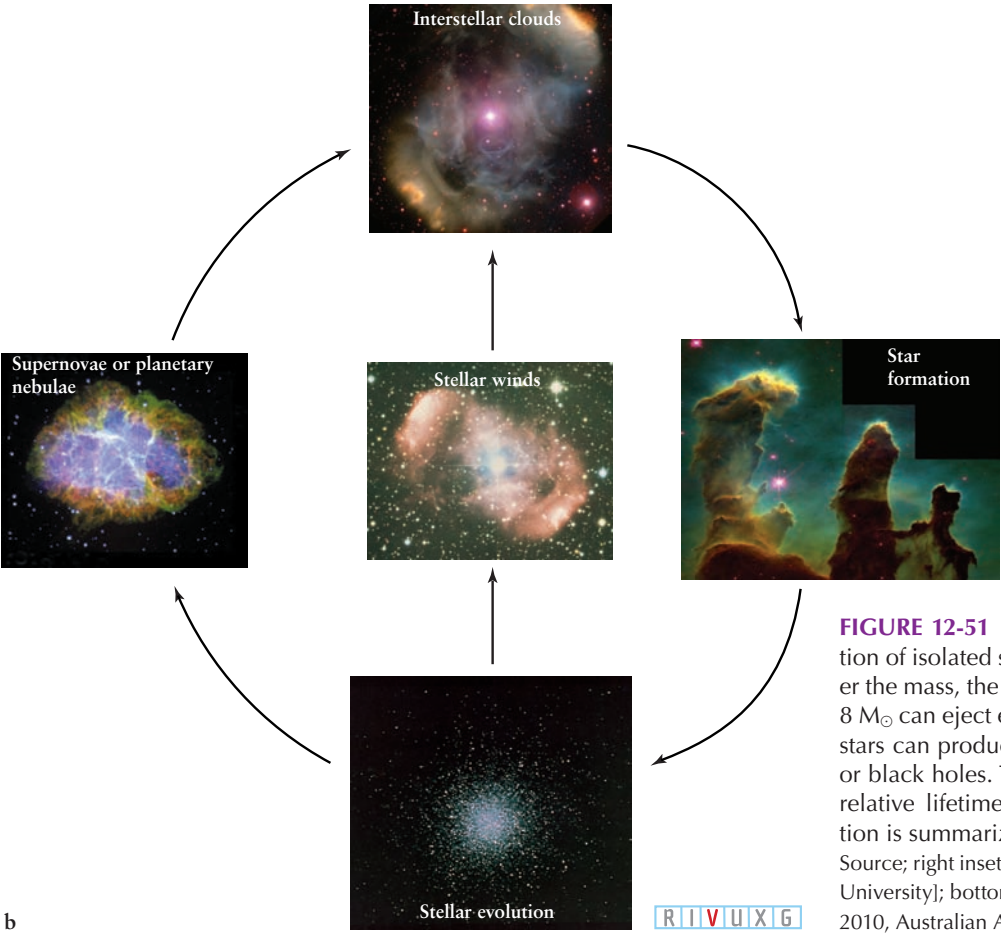
Figure 12-51 summarizes the evolutionary history of stars.

Focus Question 12-23

Why does at least one particle in the Hawking process always fall into the black hole?



a



b

FIGURE 12-51 A Summary of Stellar Evolution (a) The evolution of isolated stars depends primarily on their masses: the higher the mass, the shorter the lifetime. Stars less massive than about $8 M_{\odot}$ can eject enough mass to become white dwarfs. High-mass stars can produce Type II supernovae and become neutron stars or black holes. The horizontal (time) axis is not to scale, but the relative lifetimes are accurate. (b) The cycle of stellar evolution is summarized in this figure. (a: Imagineering; top inset: Science Source; right inset: NASA/ESA/STScI/J. Hester & P. Scowen [Arizona State University]; bottom inset: NASA; left inset: NASA; middle inset: © 1980–2010, Australian Astronomical Observatory, photograph by David Malin)

SUMMARY OF KEY IDEAS

- Stars with higher masses fuse more elements into existence than do stars with lower masses.
- Stars lose mass via stellar winds throughout their lives.

Intermediate-Mass Stars and Planetary Nebulae

- An intermediate-mass (between 0.4 and $8 M_{\odot}$) main-sequence star becomes a giant when hydrogen shell fusion begins. It becomes a horizontal-branch star when core helium fusion begins. It enters the asymptotic giant branch and becomes a supergiant when helium shell fusion starts.
- Stellar winds during the thermal pulse phase eject mass from the star's outer layers.
- The burned-out core of an intermediate-mass star becomes a dense carbon-oxygen body, called a white dwarf, with about the same diameter as that of Earth. The maximum mass of a white dwarf (the Chandrasekhar limit) is $1.4 M_{\odot}$.
- Explosive hydrogen fusion may occur in the surface layer of a white dwarf in some close binary systems, producing sudden increases in luminosity that we call novae.
- An accreting white dwarf in a close binary system can also become a Type Ia supernova when carbon fusion ignites explosively throughout such a degenerate star.

High-Mass Stars and Supernovae

- After exhausting its central supply of hydrogen and helium, the core of a high-mass (above $8 M_{\odot}$) star undergoes a sequence of other thermonuclear reactions. These reactions include carbon fusion, neon fusion, oxygen fusion, and silicon fusion. This last fusion eventually produces an iron core.
- A high-mass star dies in a supernova explosion that ejects most of the star's matter into space at very high speeds. This Type II supernova is triggered by the gravitational collapse and subsequent bounce of the doomed star's core.
- Neutrinos were detected from Supernova 1987A, which was visible to the naked eye. Its development supported theories of Type II supernovae.

Neutron Stars and Pulsars

- The core of a high-mass main-sequence star containing between $8 M_{\odot}$ and $25 M_{\odot}$ becomes a neutron star. A neutron star is a very dense stellar corpse consisting of closely packed neutrons in a sphere roughly 20 km in diameter.
- A pulsar is a rapidly rotating neutron star with a powerful magnetic field tilted relative to the star's rotation axis. The spinning field makes the neutron star a source of periodic radio and other electromagnetic pulses. Energy pours out of the polar regions of the neutron star in intense beams that sweep across the sky.
- Some X-ray sources exhibit regular pulses. These objects are believed to be neutron stars in close binary systems with ordinary stars.

- Explosive helium fusion may occur in the surface layer of a companion neutron star, producing a sudden increase in X-ray radiation, called an X-ray burster.
- If a stellar corpse is more massive than about $3 M_{\odot}$, gravitational compression overcomes neutron degeneracy and forces it to collapse further and become a black hole.
- A black hole is an object so dense that the escape velocity from it exceeds the speed of light.

The Relativity Theories

- Special relativity reveals that space and time are intimately connected and change with an observer's relative motion.
- As seen by observers moving more slowly, the faster an object moves, the slower time passes for it (time dilation) and the shorter it becomes (length contraction).
- According to general relativity, mass causes space to curve and time to slow down. These effects are significant only near large masses or compact objects.

Inside a Black Hole

- The event horizon of a black hole is a spherical boundary where the escape velocity equals the speed of light. No matter or electromagnetic radiation can escape from inside the event horizon. The distance from the center of the black hole to the event horizon is the Schwarzschild radius.
- The matter inside a black hole collapses to a singularity. The singularity for nonrotating matter is a point at the center of the black hole. For rotating matter, the singularity is a ring inside the event horizon.
- Matter inside a black hole has only three physical properties: mass, angular momentum, and electric charge.
- Nonrotating black holes are called Schwarzschild black holes. Rotating black holes are called Kerr black holes. The event horizon of a Kerr black hole is surrounded by an ergoregion in which all matter must constantly move to avoid being pulled into the black hole.
- Matter that approaches a black hole's event horizon is stretched and torn by the extreme tidal forces generated by the black hole, light from the matter is redshifted, and time slows down.
- Black holes can evaporate by the Hawking process, in which virtual particles near the black hole become real. These transformations of virtual particles into real ones decrease the mass of a black hole until, eventually, it disappears.

Evidence of Black Holes

- Observations indicate that some binary star systems harbor black holes. In such systems, gases captured by the black hole from the companion star heat up and emit detectable X-rays and jets of gas.

- Supermassive black holes exist in the centers of most galaxies. Intermediate-mass black holes appear to exist in globular clusters of stars. Very low mass (primordial) black holes may have formed at the beginning of the universe.

Gamma-Ray Bursts

- Gamma-ray bursts are events believed to be caused by some supernovae and by the collisions of dense objects, such as neutron stars or black holes. Some occur in the Milky Way and nearby galaxies, whereas many others occur billions of light-years away from Earth.
- Typical gamma-ray bursts occur for a few tens of seconds and emit more energy than the Sun will radiate over its entire 10-billion-year lifetime.

WHAT DID YOU THINK?

1. *Will the Sun someday cease to shine brightly? If so, how will this occur?* Yes. The Sun will shed matter as a planetary nebula in about 6 billion years and then cease nuclear fusion. Its remnant white dwarf will dim over the succeeding billions of years.
2. *What is a nova? How does it differ from a supernova?* A nova is a relatively gentle explosion of hydrogen gas on the surface of a white dwarf in a binary star system. Supernovae, on the other hand, are explosions that cause the nearly complete destruction of massive stars.
3. *What are the origins of the carbon, silicon, oxygen, iron, uranium, and other heavy elements on Earth?* These elements are created during stellar evolution, by supernovae, and by colliding neutron stars.
4. *What are cosmic rays? Where do they come from?* Cosmic rays are high-speed particles (mostly hydrogen and other atomic nuclei) in space. Many of them are created as supernova remnants collide with preexisting interstellar gas.
5. *What is a pulsar?* A pulsar is a rotating neutron star in which the magnetic field's axis does not coincide with the rotation axis. The beam of radiation it emits periodically sweeps across our region of space.
6. *Are black holes empty holes in space? If not, what are they?* No. Black holes contain highly compressed matter—they are not empty.
7. *Does a black hole have a solid surface? If not, what is at its surface?* No. The surface of a black hole, called the *event horizon*, is empty space. No stationary matter exists there.
8. *What power or force enables black holes to draw things into themselves?* The only force that pulls things in is the gravitational attraction of the matter and energy in the black hole.

9. *How close to a black hole do you have to be for its special effects to be apparent?* An object within about 100 times the Schwarzschild radius from a black hole will begin to feel its special effects, like time dilation.

10. *Can you use black holes to travel to different places in the universe?* No. Most astronomers believe that the wormholes predicted by general relativity do not exist.

11. *Do black holes last forever? If not, what happens to them?* No. Black holes evaporate.

Review Questions

1. Will the Sun shed most of its mass, and, if so, what is that event called?
 - a. yes, as a planetary nebula
 - b. yes, as a supernova
 - c. yes, as a white dwarf
 - d. yes, as a neutron blast
 - e. no
2. A white dwarf is composed of primarily
 - a. neutrons
 - b. hydrogen and helium
 - c. iron
 - d. cosmic rays
 - e. carbon and oxygen
3. What prevents a neutron star from collapsing?
 - a. hydrogen fusion
 - b. friction
 - c. electron degeneracy pressure
 - d. neutron degeneracy pressure
 - e. helium fusion
4. What is the difference between a giant star and a supergiant star?
5. Why is knowing the temperature in a star's core so important in determining which nuclear reactions can occur there?
6. What determines the temperature in the core of a star?
7. What is a planetary nebula, and how does it form?
8. What is the Chandrasekhar limit?
9. What is a neutron star?
10. Compare a white dwarf and a neutron star. Which of these stellar corpses is more common? Why?
11. What is the mass range of neutron stars?
12. Why have searches for supernova remnants at visible wavelengths been less fruitful than searches at other wavelengths?
13. Why do astronomers believe that pulsars are rapidly rotating neutron stars?
14. What is the difference between Type Ia and Type II supernovae?

15. Compare a nova with a Type Ia supernova. What do they have in common? How are they different?
16. Describe what X-ray pulsars, pulsating X-ray sources, and X-ray bursters have in common. How are they different manifestations of the same type of astronomical object?
17. Which property, if any, of normal matter ceases to exist in a black hole?
 - a. mass
 - b. chemical composition
 - c. angular momentum
 - d. electric charge
 - e. All of these properties exist in a black hole.
18. Supermassive black holes are found in which of the following locations?
 - a. in the centers of galaxies
 - b. in globular clusters
 - c. in open (or galactic) clusters
 - d. between galaxies
 - e. in orbit with a single star
19. Which feature is found with Kerr black holes but not Schwarzschild black holes?
 - a. a singularity
 - b. an event horizon
 - c. gravitational redshift of photons outside of the black hole
 - d. an ergoregion
 - e. warping of nearby spacetime
20. Under what conditions do all outward pressures on a collapsing star fail to stop its inward motion?
21. In what way is a black hole blacker than black ink or a black piece of paper?
22. If the Sun suddenly became a black hole, how would Earth's orbit be affected?
23. What are the differences between rotating and nonrotating black holes?
24. Why are all of the observed stellar-remnant black-hole candidates members of close binary systems?
25. If light cannot escape from a black hole, how can we detect X-rays from such an object?

Advanced Questions

26. What prevents thermonuclear reactions from occurring at the center of an isolated white dwarf? If no thermonuclear reactions take place in its core, why doesn't that body collapse?
27. Suppose you wanted to determine the age of a planetary nebula. What observations would you make, and how would you use the resulting data?
28. Why is the rate of expansion of the gas shell in a planetary nebula often not uniform in all directions?
29. To determine the period of a pulsar accurately, astronomers must take Earth's orbital motion around the Sun into account. Explain why.
30. If more massive stars evolve and die before less massive ones, why do some observed black-hole candidates have lower masses than their stellar companions?
31. Under what circumstances might a neutron star in a binary star system become a black hole?
32. Which type of black hole, nonrotating or rotating, do science fiction writers use (implicitly) in sending spaceships from one place to another through the hole? Why would the other type not be suitable?

Discussion Questions

33. Suppose that you discover a small glowing disk of light while searching the sky with a telescope. How would you determine whether this object was a planetary nebula? What else could this object be?
34. Describe how astronomers can determine whether a supernova at a known distance is Type Ia or Type II, assuming that they can see the supernova from the time it begins to brighten. There are at least two valid answers to this question.
35. Describe how astronomers can determine whether a supernova at an unknown distance is Type Ia or Type II.

Got It?

36. How will the Sun end its "life cycle"? Will there be anything left of it after fusion in it ceases?
37. Where was the iron in your blood formed?
38. What are cosmic rays?
39. When a star leaves the main sequence, where is the energy being generated that pushes its outer layers outward?
40. Are black holes *completely* black (that is, emitting or scattering no light) and are they holes (in the sense of being completely empty)? Explain.
41. How does the mass of a young black hole compare to the mass of the neutron star from which it just formed?
42. Are black holes like giant vacuum cleaners in space that are eventually going to suck up everything in the universe? Explain your reasoning.
43. Which of the following statements most accurately describes the fate of a black hole?
 - a. It will remain unchanged as long as the universe exists.
 - b. It will eventually suck in everything else in the universe.
 - c. It will evaporate.
 - d. It will eventually start shining due to fusion in its core.
 - e. It will eventually start pulsating by periodically growing larger and then smaller.

This page intentionally left blank



CHAPTER 13

R I V U X G The Sombrero Galaxy (also designated M104, the 104th galaxy listed in the Messier catalog) is 29 million ly from us. By combining infrared, optical, and X-ray observations, we can gain insights into its disk of stars, gas, and dust, along with its central region, and the hot gas surrounding it. (X-ray: NASA/UMass/Q. D. Wang et al.; Optical: NASA/STScI/AURA/Hubble Heritage; Infrared: NASA/JPL-Caltech/Univ. AZ/R. Kennicutt/SINGS Team)

The Galaxies

WHAT DO YOU THINK?

- 1 What is the shape of the Milky Way Galaxy?
- 2 Where is our solar system located in the Milky Way Galaxy?
- 3 Is the Sun moving through the Milky Way Galaxy and, if so, about how fast?
- 4 Are most of the stars in spiral galaxies located in their spiral arms?
- 5 Do all galaxies have spiral arms?
- 6 Are galaxies isolated objects?
- 7 Is the universe contracting, unchanging in size, or expanding?
- 8 What does “quasar” stand for?
- 9 What do quasars look like?
- 10 Where do quasars get their energy?

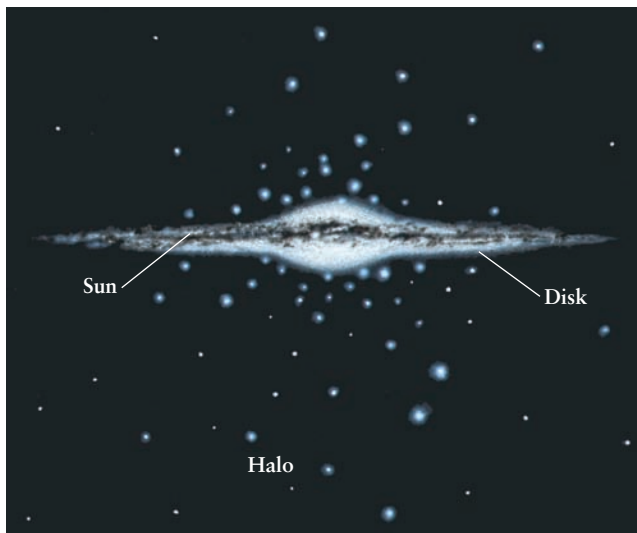
Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

Centuries of observations have firmly established that our solar system is part of an enormous assemblage of hundreds of billions of stars, along with gas, dust, and other matter, all held together by their mutual gravitational attraction. This is our **Milky Way Galaxy**. Most of the stars in our Galaxy are located in a disk that looks from the edge like a flying saucer in an old science fiction movie (**Figure 13-1a**). Spiral arms swirl out from the ends of the bar and, within these arms, new stars form from the debris of earlier generations of stars. The Galaxy's remaining stars are located in a two-shell, spherical halo that surrounds the disk. (*Note:* The word “Galaxy” when used alone, as here, is capitalized only when it refers to our Milky Way.)

After considering our Galaxy, we explore the other types of galaxies, the groupings of galaxies called clusters and superclusters of galaxies, and the nature of the space that connects them. We end this chapter by examining the supermassive black holes that lie at the centers of most galaxies, and the effects these black holes have on the surrounding matter.

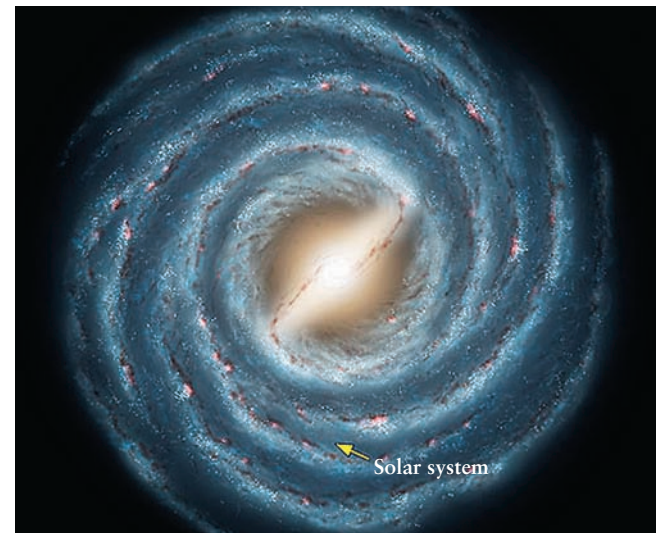
In this chapter you will discover

- the Milky Way Galaxy—billions of stars along with stellar remnants, gas, dust, and dark matter that are all bound together by their mutual gravitational attraction
- the structure of our Milky Way Galaxy and Earth's location in it
- that there is a supermassive black hole at the center of our Galaxy
- how galaxies are categorized by their shapes
- that galaxies are found in clusters that contain huge amounts of dark matter and why clusters of galaxies form in superclusters
- how some galaxies merge and others devour their neighbors
- that the universe is changing size
- bright and unusual objects called active galaxies and distant, luminous quasars
- supermassive black holes that serve as central engines for radio galaxies, quasars, Seyfert galaxies, and the extremely powerful BL Lac objects



a

FIGURE 13-1 Schematic Diagrams of the Milky Way (a) This edge-on view shows the Milky Way's disk, containing most of its stars, gas, and dust, and its halo, containing many old stars. Individual stars in the halo are too dim to be visible on this scale, so the bright regions in the halo represent clusters of stars.



b

(b) Our Galaxy has two major arms and several shorter arm segments, all spiraling out from the ends of a bar of stars and gas that passes through the Galaxy's center. The bar's existence and the presence of two major arms were confirmed by the Spitzer Space Telescope. (b: NASA/JPL-Caltech/R. Hurt [SSC])

THE MILKY WAY

Prior to the twentieth century, astronomers did not know the large-scale distribution of stars and other matter in the universe. Throughout history, most people, including many astronomers, believed that the Milky Way contains all of the stars in the cosmos. In other words, they thought that the “Galaxy” and the “universe” were the same thing. We begin the study of galaxies by learning how that belief changed.

13-1 Studies of Cepheid variable stars revealed that the Milky Way is only one of many galaxies

The concept that our Galaxy is but one of many was put forth in 1755, when the German philosopher Immanuel Kant suggested that vast collections of stars lie far beyond the confines of the Milky Way. Less than a century later, the Irish astronomer William Parsons observed the structure of some of those “island universes” proposed by Kant. However, most astronomers of Parsons’s day did not agree with the notion of island universes outside of our Galaxy. They thought that the Milky Way contained all stars in the universe—indeed, that the Milky Way *was* the universe.

In April 1920, a formal discussion, now known as the Shapley–Curtis debate, was held at the National Academy of Sciences in Washington, D.C. Harlow Shapley argued that the spiral nebulae are relatively small, nearby objects scattered around our Galaxy. Heber D. Curtis championed the island universe theory, arguing that each of these spiral nebulae is a separate rotating system of stars, much like our own Galaxy. While the Shapley–Curtis debate focused scientific attention on the size of the universe, nothing was decided, because no one had any firm evidence to demonstrate exactly how far away the spiral nebulae were. Astronomers desperately needed to devise a way to measure the distances to the nebulae. A young teacher and former basketball player originally from Missouri, who moved to Chicago to study astronomy, finally achieved this goal. His name was Edwin Hubble.

In 1923, Hubble took a historic photograph of M31, then called the Andromeda Nebula. It was one of the spiral nebulae around which controversy raged. On the photographic plate he discovered what first appeared to be a nova. Referring to previous plates of that region, he soon realized that the object was actually a Cepheid variable star. As we saw in Section 11-12, these pulsating stars vary in brightness periodically. Further scrutiny over the next several months revealed many other Cepheids. **Figure 13-2** shows a Cepheid in the galaxy M100, at different stages of brightness.

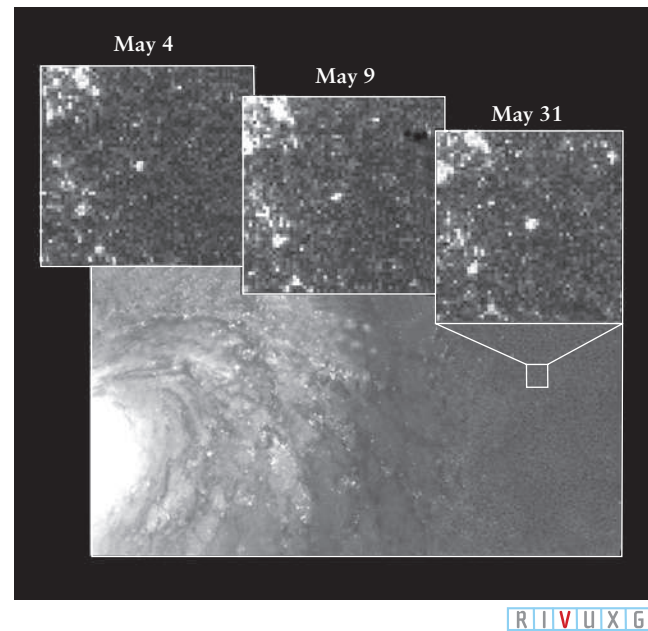


FIGURE 13-2 A Cepheid Variable Star in Galaxy M100

One of the most reliable ways to determine the distance to moderately remote galaxies is to locate Cepheid variable stars in them, as discussed in the text. The distance of 50 million ly (15.2 Mpc) from Earth to the galaxy M100 in the constellation Coma Berenices was determined using Cepheids. **Insets:** The Cepheid in this view, one of 20 located to date in M100, is shown at different stages in its brightness cycle, which recurs over several weeks. (Dr. Wendy L. Freedman, Observatories of the Carnegie Institution of Washington; NASA)

Only a decade before, in 1912, the American astronomer Henrietta Leavitt had published an important study of Cepheid variables. Leavitt studied many Cepheids in the Small Magellanic Cloud, then also believed to be a nebula, but now known to be a galaxy passing very close to the Milky Way. Leavitt’s study led her to the period–luminosity relation for Type I Cepheids (see Section 11-13). Leavitt established that a direct relationship exists between a Cepheid’s luminosity (or absolute

Insight Into Science

Room for Debate Lacking definitive data, competent scientists can develop and believe strikingly different explanations for the same observations. At the time of the Shapley–Curtis debate, the observations allowed both points of view. It was only with the advent of the distance measurement technique used by Hubble that the spiral nebulae were definitely shown to lie outside of the Galaxy.

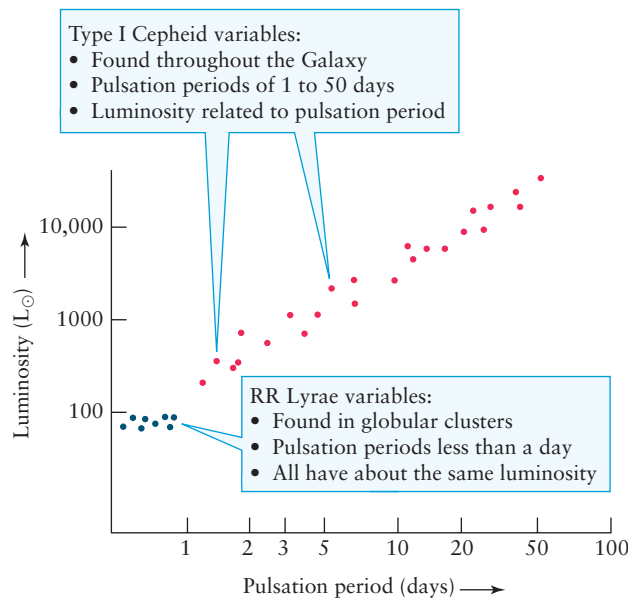


FIGURE 13-3 The Period–Luminosity Relation This graph shows the relationship between the periods and average luminosities of classical (Type I) Cepheid variables and the closely related RR Lyrae stars (discussed in Chapter 11). Each dot represents a Cepheid or RR Lyrae whose luminosity and period have been measured.

magnitude) and its period of oscillation, which we saw in Figure 11-25 and which is presented in more detail in **Figure 13-3**. By observing the star’s period and apparent magnitude, its distance can be calculated.

Thanks to this work, Hubble knew that he could use the characteristics of the fluctuating light of Cepheid variable stars to help him calculate the distance to M31 and put the Shapley–Curtis debate to rest once and for all. His observations of Type I Cepheids led him to determine that M31 is some 2.2 million ly *beyond* the Milky Way. This distance proves that M31 is not a young star and planet-forming disk of gas and dust, or an open or globular cluster in our Galaxy, but rather an enormous separate stellar system—a separate galaxy. M31, now called the *Andromeda Galaxy*, is the most distant object in the universe that can be seen with the naked eye. Similar calculations have been done for all galaxies in which Cepheids can be observed. The distances to even more remote galaxies have been determined from observations of Type Ia supernovae in them.

Hubble’s results, which he presented at the end of 1924, did settle the Shapley–Curtis debate. The universe was recognized to be far larger and populated with far bigger objects than most astronomers had imagined. Hubble had discovered the realm of

the galaxies. Today, we know that the universe contains myriad galaxies, of which the Milky Way is just one. Like the Milky Way, each **galaxy** is a grouping of millions, billions, or even trillions of stars, along with gas, dust, and matter in other forms, all gravitationally bound together.

We now apply Hubble’s method to find Earth’s place in the Milky Way Galaxy. Recall that the Sun’s proximity to us makes it our best-understood star. It might seem that the nearness of the stars and clouds in the Milky Way would make it the best-understood galaxy. However, the clouds of gas and dust that surround the solar system make it very challenging for astronomers to survey completely the distant parts of the Galaxy, which are only now coming into focus.

13-2 Cepheid variables help us locate our Galaxy’s center

Because the band of the Milky Way completely encircles us, astronomers long ago suspected that the Sun and all of the stars that we see are part of it. In the 1780s, William Herschel took the first steps toward mapping our Galaxy’s structure. He attempted to deduce the Sun’s location in the Galaxy by counting the number of stars in 683 regions of the sky. He reasoned that the greatest density of stars should be seen toward the Galaxy’s center and a lesser density seen toward the edge. However, Herschel found roughly the same density of stars all along the Milky Way. He therefore concluded that we are at the center of the Galaxy.

Herschel was wrong: Earth has no privileged place in the Milky Way. As we will now see, the Sun is about 26,000 ly (8000 pc) from the Galaxy’s center, the **galactic nucleus**. Herschel’s physical understanding of the cosmos was incomplete, so he misinterpreted his observations and thus came to an incorrect conclusion.

While studying star clusters in the 1930s, Robert Trumpler discovered the reason for Herschel’s mistake. Herschel did not know about interstellar gas and dust, which affected his counts of the stars. Trumpler noticed that remote star clusters appear dimmer than would be expected just from their distance alone. Something must

Insight Into Science

A Little Knowledge Incomplete information often leads to incorrect interpretation of data and, therefore, to incorrect conclusions. Herschel’s lack of knowledge about the matter in interstellar space prevented him from correctly interpreting the distribution of stars that surround Earth and, thus, led to his inaccurate conclusion about the position of the Sun within the Galaxy.

Focus Question 13-1

Referring to Figure 13-3, how would the brightness of a Cepheid variable with peak luminosity of $1000 L_{\odot}$ change if it were observed every 5 days?



R11VUX16

FIGURE 13-4 Our Galaxy This wide-angle photograph spans half the Milky Way. The center of the Galaxy is in the constellation Sagittarius, in the middle of this photograph. The

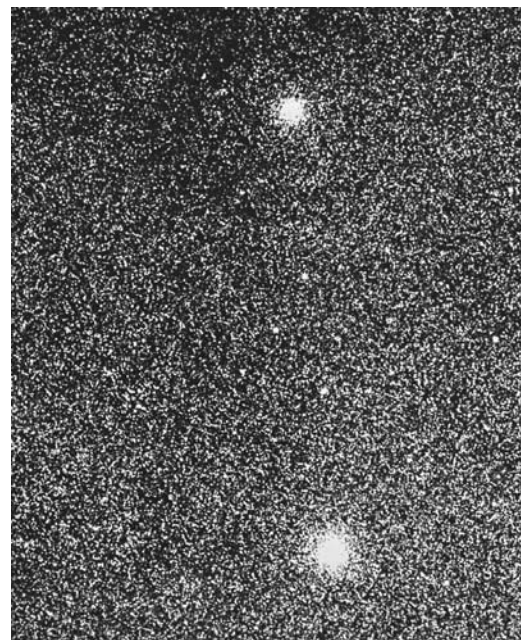
dark lines and blotches are caused by hundreds of interstellar clouds of gas and dust that obscure the light from background stars rather than by a lack of stars. (Eckhard Slawik/Science Source)

be blocking starlight on its way toward Earth. He correctly concluded that interstellar space is not a perfect vacuum. Instead, it contains dust that absorbs light from distant stars. Great patches of this dust are clearly visible in wide-angle photographs (Figure 13-4). Like the stars, this dust is concentrated in the plane of the Galaxy.

This interstellar dust almost completely obscures from view visible light emanating from the center of our Galaxy. Visual photons from there are mostly absorbed or scattered before they reach us. Therefore, Herschel was seeing only nearby stars, and he measured apparent magnitudes that were dimmer than they would have been had there been no interstellar dust. Without adjusting for its effects, he concluded that the stars were farther away than they really are. He also had no idea of the true size of the Galaxy and could not see the vast number of stars located in the general direction of the galactic center that are hidden by the dust.

Because interstellar dust is concentrated in the plane of our Galaxy's disk, the absorption of starlight is strongest in those parts of the sky covered by the Milky Way. Above or below the plane of the Galaxy, our view is relatively unobscured. Knowledge of our true position in the Galaxy eventually came from observations of globular clusters (see Section 11-14). Shapley used the period–luminosity relation for Cepheid variable stars to determine the distances to the then-known 93 globular clusters in the sky. (More than 150 are known today.) From their directions and distances, he mapped out the distribution of these clusters in three-dimensional space. By 1917, Shapley had discovered that the globular clusters are located in a spherical distribution centered not on Earth but on a point in the Milky Way toward the constellation Sagittarius. Figure 13-5 shows two globular clusters in a relatively clear part of the sky in that di-

rection. Shapley then made a bold conjecture: *The globular clusters orbit the center of the Milky Way, which is located in Sagittarius*. His pioneering research has since been observationally verified. Earth is not at the center of the Galaxy.



R11VUX16

FIGURE 13-5 A View Toward the Galactic Center More than a million stars in the disk of our Galaxy fill this view, which covers a relatively clear window just 4° south of the galactic nucleus in Sagittarius. Beyond the disk stars you can see two prominent globular clusters. Although most regions of the sky toward Sagittarius are thick with dust, very little obscuring matter appears in this tiny section of the sky. (NOAO/AURA/NSF)

13-3 Nonvisible observations help map the galactic disk

To see into the dust-filled plane of the Milky Way, astronomers use radio wave, infrared, X-ray, and gamma-ray telescopes (see Figure 3-37). These wavelengths are scattered much less by the interstellar gas and dust located throughout the Galaxy's disk than are visible or ultraviolet wavelengths. Observations of the distant parts of the Galaxy were first made using radio telescopes. Radio waves penetrate Earth's atmosphere, so we can observe them anywhere that we can build a radio telescope.

Focus Question 13-2

What situation(s) on Earth are analogous to the obscuration caused by interstellar gas and dust?

(Recall that infrared observations must be made at high altitudes or in space and that X-ray and gamma-ray observations are almost always made from space.)

Detecting the radio emission directly from interstellar hydrogen—by far the most abundant element in the universe—is a primary means of mapping the Galaxy. Unfortunately, the major transitions of electrons in the hydrogen atom (see Figure 3-51) produce photons at ultraviolet and visible wavelengths that do not penetrate the interstellar medium. How, then, can radio telescopes directly detect all of this hydrogen? The answer lies in atomic physics.

In addition to mass and charge, particles such as protons and electrons possess a tiny amount of angular momentum, commonly called **spin**. According to the laws of quantum mechanics, the electron and proton in a hydrogen atom can spin only in either parallel or opposite directions (Figure 13-6); they can have no other spin orientations. If the electron in a hydrogen atom flips from one orientation to the other, the atom must gain or lose a tiny amount of energy. In particular, when flipping from parallel to opposite spins, the atom simultaneously emits a low-energy radio photon whose wavelength is 21 cm. This flip happens rarely in each atom, so it is only because the Galaxy has vast quantities of interstellar hydrogen gas that it can be detected at all. In 1951, a team of astronomers first succeeded in detecting the faint hiss of 21-cm radio static from spin flips.

The detection of **21-cm radio radiation** was a major breakthrough in mapping the **disk** of the Galaxy. To see why, suppose that you aim your radio telescope across the Galaxy, as sketched in Figure 13-7. Your radio receiver picks up 21-cm emission from hydrogen clouds at points 1, 2, 3, and 4. However, the radio waves from these various clouds are Doppler shifted (see Section 3-2) by slightly different amounts because they have different radial velocities (motion toward or away from Earth). Because these radio waves from gas clouds in different parts of the Galaxy arrive at our radio telescopes with slightly different wavelengths as a result of being Doppler shifted, it is possible to identify which radio signals come from which gas clouds and thus to produce an initial map of the Galaxy, such as that shown in Figure 13-8a.

Our radio map reveals numerous arched lanes of neutral hydrogen gas. If this were the overall structure

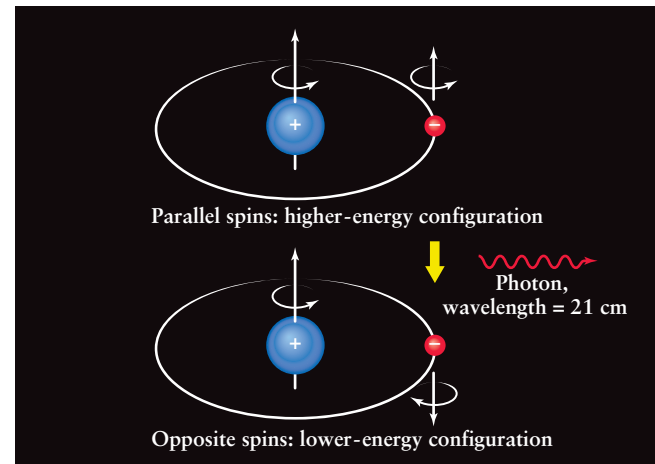


FIGURE 13-6 Electron Spin and the Hydrogen Atom Due to their spins, electrons and protons both act as tiny magnets. When an electron and the proton it orbits are spinning in the same direction, their energy is higher than when they are spinning in opposite directions. When the electron flips from the higher-energy to the lower-energy configuration, the atom loses a tiny amount of energy that is radiated as a radio photon with a wavelength of 21 cm (8.3 in).

of the Galaxy, then the Milky Way would appear unlike any other observed galaxy (for comparison see the radio image of the barred spiral galaxy M83 in Figure 13-9a and Sections 13-8 through 13-12 for typical visible light images of other galaxies). Indeed, the other disk-shaped galaxies we observe have spiral arms (Figure 13-9b). Dif-

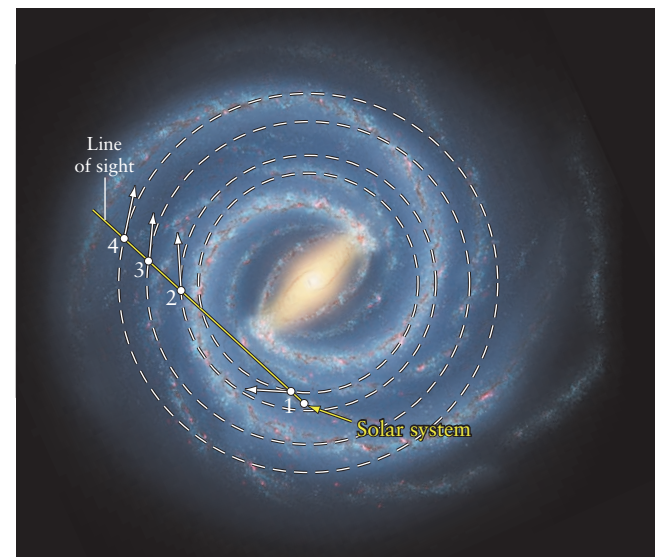


FIGURE 13-7 A Technique for Mapping the Galaxy Hydrogen clouds at different locations along our line of sight are moving around the center of the Galaxy at different speeds. The component of their motion away from us varies with their distance from the solar system. Radio waves from the various gas clouds, therefore, exhibit slightly different Doppler shifts, permitting astronomers to sort out the gas clouds and map the Galaxy. (NASA/JPL-Caltech/ESO/R. Hurt)

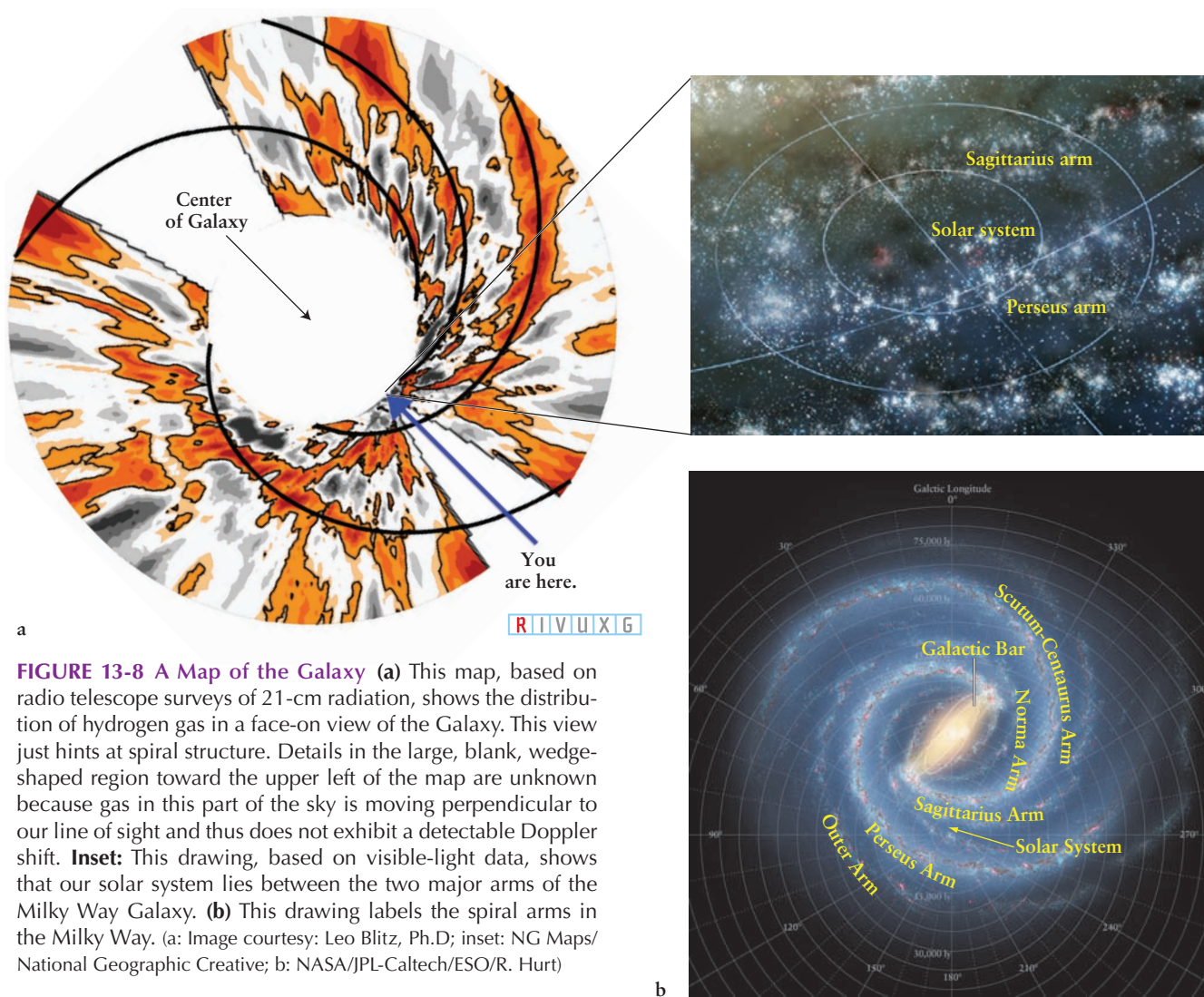
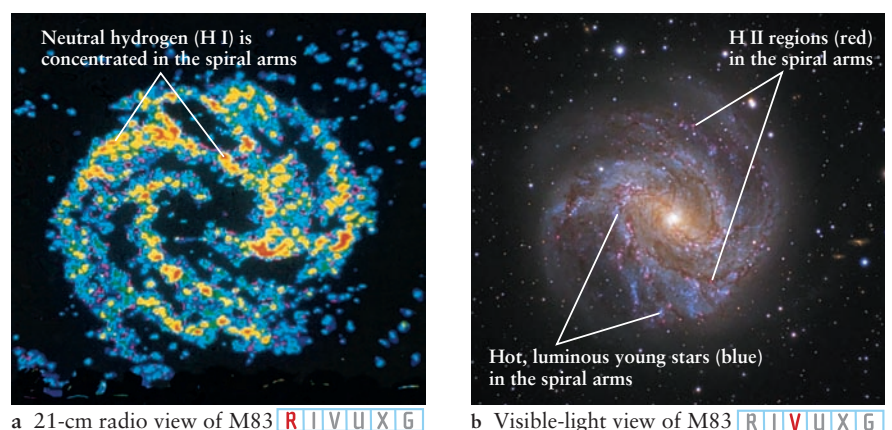


FIGURE 13-8 A Map of the Galaxy (a) This map, based on radio telescope surveys of 21-cm radiation, shows the distribution of hydrogen gas in a face-on view of the Galaxy. This view just hints at spiral structure. Details in the large, blank, wedge-shaped region toward the upper left of the map are unknown because gas in this part of the sky is moving perpendicular to our line of sight and thus does not exhibit a detectable Doppler shift. **Inset:** This drawing, based on visible-light data, shows that our solar system lies between the two major arms of the Milky Way Galaxy. (b) This drawing labels the spiral arms in the Milky Way. (a: Image courtesy: Leo Blitz, Ph.D; inset: NG Maps/ National Geographic Creative; b: NASA/JPL-Caltech/ESO/R. Hurt)

ferent observations were needed to improve our understanding of our Galaxy's disk features. Note that photographs of a barred spiral galaxy (Figure 13-9b) show arms outlined by “spiral tracers”—bright, Population I stars and emission nebulae. As we saw in Chapter 11, these features indicate active star formation. If the Milky Way is spiral, then another useful way to further chart its structure and show that it has **spiral arms** is to map the locations of star-forming complexes filled with H II regions, giant molecular clouds, and massive, hot, young stars in groups, called *OB associations*.

1 Dust absorption limits the range of visual observations in the plane of the Galaxy to less than 24,000 ly from Earth. However, astronomers can use visible observations of nearby bright OB associations and associated H II regions to plot the spiral arms near the Sun (see Figure 13-8a inset). Radio observations of hydrogen and carbon monoxide molecules (discussed in Section 11-1) have been used to chart more remote star-forming regions of the Galaxy. Taken together, all of these observations indicate that our Galaxy has about 200 billion stars located in and between two major spiral arms and several minor arms (see

Figure 13-8b). (Each arm is named after the constellation in which it is centered, as seen from Earth.) We will explore the origin of the spiral arms in Section 13-8, when we present more evidence about the cause of spiral structure from



a 21-cm radio view of M83 **R I V I U X I G** b Visible-light view of M83 **R I V I U X I G**

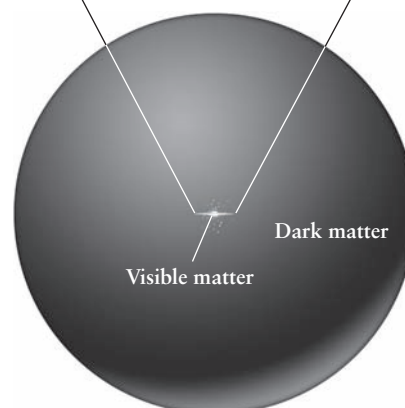
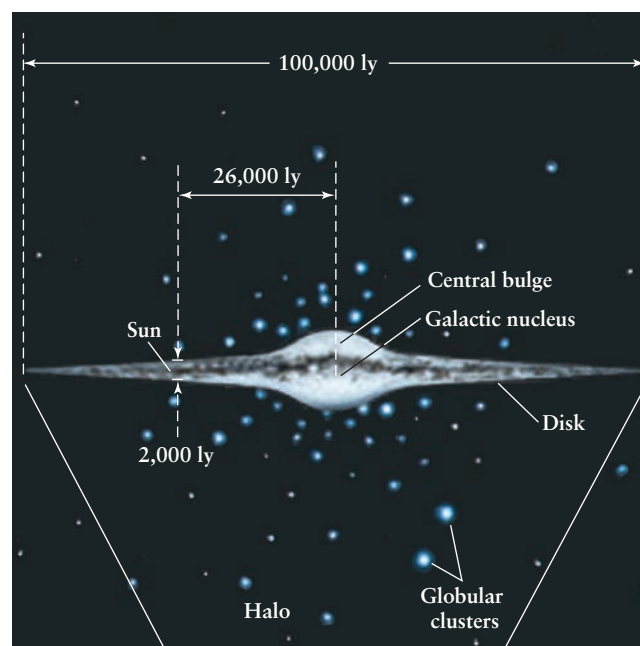
FIGURE 13-9 Two Views of a Barred Spiral Galaxy The galaxy M83 is in the southern constellation of Centaurus, about 12 million ly from Earth. (a) A radio view at 21-cm wavelength shows the emission from neutral hydrogen gas. (b) Note that the spiral arms are more clearly demarcated by visible stars and H II regions than by 21-cm radio emission. (a: Atlas Image courtesy of 2MASS/UMass/IPAC-Caltech/NASA/NSF; b: Robert Gendler and Stephane Guisard)

FIGURE 13-10 Our Galaxy As seen from the side, three major visible components of our Galaxy are a thin disk, a central bulge, and a two-part halo system. As noted earlier, there is also a central bar. The visible Galaxy's diameter is about 100,000 ly, and the Sun is about 26,000 ly from the galactic center. The disk contains gas and dust along with Population I (young, metal-rich) stars. The halo is composed almost exclusively of Population II (old, metal-poor) stars. **Inset:** The visible matter in our Galaxy fills only a small volume compared to the distribution of dark matter, whose composition is presently unknown. Dark matter's presence is felt by its gravitational effect on visible matter.

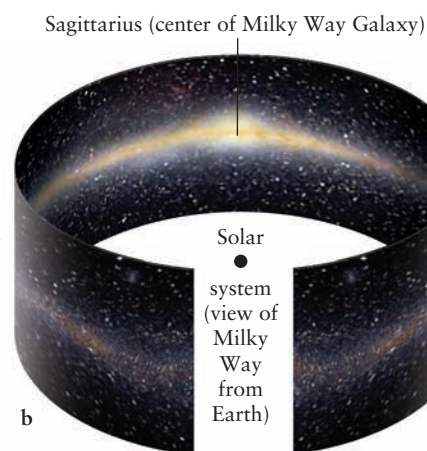
observations of other galaxies. Recent observations by the Spitzer Space Telescope confirm previous observations that a bar of stars and gas crosses the center of the Galaxy, also shown in Figure 13-8b. The stars, gas, and dust in the bar move down one side of it and then reverse direction and move back along the other side.

2 The observable disk of our Galaxy is about 100,000 ly in diameter and about 2000 ly thick (Figure 13-10). Its two major spiral arms are designated the Perseus Arm and the Scutum-Centaurus Arm (see Figure 13-8b). There are also several less pronounced arms, including the Sagittarius Arm, the Norma Arm, the Local Arm, and the Outer Arm. The solar system is located in the Local Arm (also called the Orion Spur), which appears to branch off the Perseus Arm. On the side toward the galactic center is the Sagittarius Arm, which stargazers in the northern hemisphere see in the summer when they look at the portion of the Milky Way stretching across Scorpius and Sagittarius (see Figure 13-4). Directed away from the galactic center is the Perseus Arm, which is visible in the northern hemisphere in the winter.

As noted above, observations reveal that the Galaxy's arms spiral out from a bar of stars, gas, and dust (Figure 13-8b) running through a flattened sphere of stars, called the **central bulge** (also called the *nuclear bulge*), that is about 12,000 ly in diameter. The central bulge is also seen in Figure 13-11, a wide-angle infrared image of the



a



b

FIGURE 13-11 Infrared View of the Milky Way (a) Taken by the COBE satellite in 1997, this infrared image shows the disk and central bulge of our Galaxy, as they would be seen from outside the Galaxy. Most of the sources scattered above and below the disk are nearby stars. Stars appear white,

whereas interstellar dust appears orange. Note that the dust that obscures light from more distant stars in Figure 13-4 is quite bright in this infrared image. (b) Because we are embedded in it, the Galaxy appears wrapped around us. (The COBE Project, DIRBE, NASA)

Galaxy taken by the COBE satellite. The central bulge is centered on the galactic nucleus 26,000 ly away from us.

13-4 The galactic nucleus is an active, crowded place

If you lived on a planet near the center of the Galaxy, which is called the galactic nucleus, you would see a million stars as bright as Betelgeuse appears from Earth today. The total intensity of starlight from all those near-

by stars would be equivalent to 200 of our full Moons. Night would never really fall. Stranger still, the Galaxy around you would be filled with intense activity.

Figure 13-12 shows three infrared views that look toward the nucleus of the Galaxy. Figure 13-12a is a wide-angle view covering a 50° segment of the Milky Way through Sagittarius and Scorpius. The prominent band across this image is a thin layer of dust in the plane of the Galaxy. The numerous knots and blobs along the dust layer are interstellar clouds heated by young O and

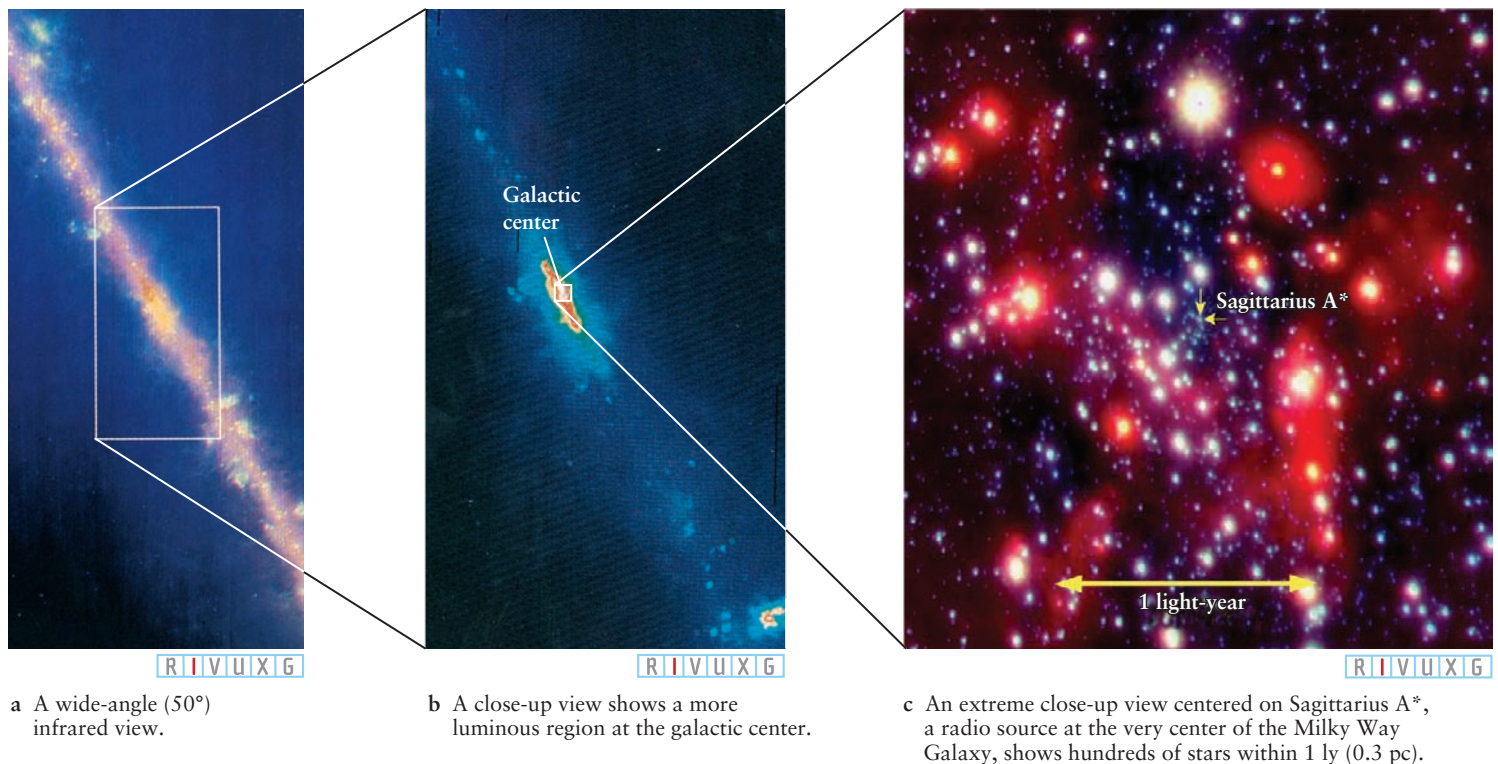
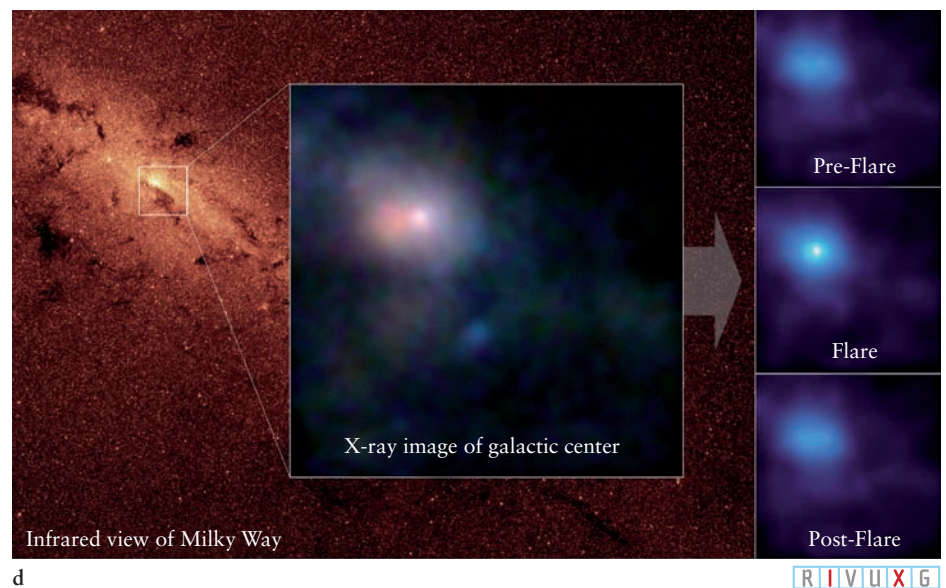


FIGURE 13-12 The Galactic Center (a) This wide-angle view at infrared wavelengths shows a 50° segment of the Milky Way centered on the nucleus of the Galaxy. Black represents the dimmest regions of infrared emission, with blue the next strongest, followed by yellow and red; white represents the strongest emission. The prominent band diagonally across this photograph is a layer of dust in the plane of the Galaxy. Numerous knots and blobs along the plane of the Galaxy are interstellar clouds of gas and dust heated by nearby stars. (b) This close-up infrared view of the galactic center covers the area outlined by the white rectangle in (a). (c) This infrared image shows about 300 of the brightest stars less than 1 ly from Sagittarius A*, which is at the center of the picture. The distribution of stars and their observed motions around the galactic center imply a very high density (about a million solar masses per cubic light-year) of less luminous stars. (d) An X-ray flare from a Sagittarius A* image in 2012, by NuSTAR (Nuclear Spectroscopic Telescope Array). (a, b: NASA; c: R. Schodel et al., MPE/ESO; d: NASA/JPL-Caltech)



B stars. Figure 13-12b is an IRAS (Infrared Astronomical Satellite) view of the galactic center. Numerous streamers of dust (blue) surround it. The strongest infrared emission (white) comes from **Sagittarius A** (often abbreviated Sgr A), which is also a grouping of several powerful sources of radio waves. One of these sources, called Sagittarius A* (pronounced “A-star”), is believed to be the galactic nucleus. Figure 13-12c shows stars within 1 ly of Sagittarius A*, with resolution of 0.02 ly.

Nonoptical observations add more to the picture of the galactic center. Some of the most detailed radio images of it come from the Very Large Array (VLA). **Figure 13-13a** is a wide-angle view of Sagittarius A and surrounding features, including arcs of gas, at least three supernovae remnants, and localized radio sources. Huge filaments, such as the one labeled “Arc,” lie perpendicular to the plane of the Galaxy and stretch 200 ly northward of the galactic

disk, then abruptly arch southward toward Sagittarius A. The orderly arrangement of these filaments suggests that a magnetic field is controlling the distribution and flow of ionized gas, just as magnetic fields on the Sun funnel such gas to create solar prominences. This resulting radio emission, produced by high-speed electrons that spiral around these magnetic fields, is called **synchrotron radiation**. Despite its small size, Sagittarius A is one of the brightest sources of synchrotron radiation in the entire sky.

X-ray observations taken by the Chandra telescope in 2004 reveal that the galactic nucleus is also bathed in ultrahot gas with a temperature of 100 million K. There is growing evidence that this gas is more than a single outburst, as flares of X-rays from gas heated to this temperature have been seen erupting from Sagittarius A* in 2001 and 2012 (Figure 13-12d). This is consistent with the fact that such hot gas needs frequent reheating

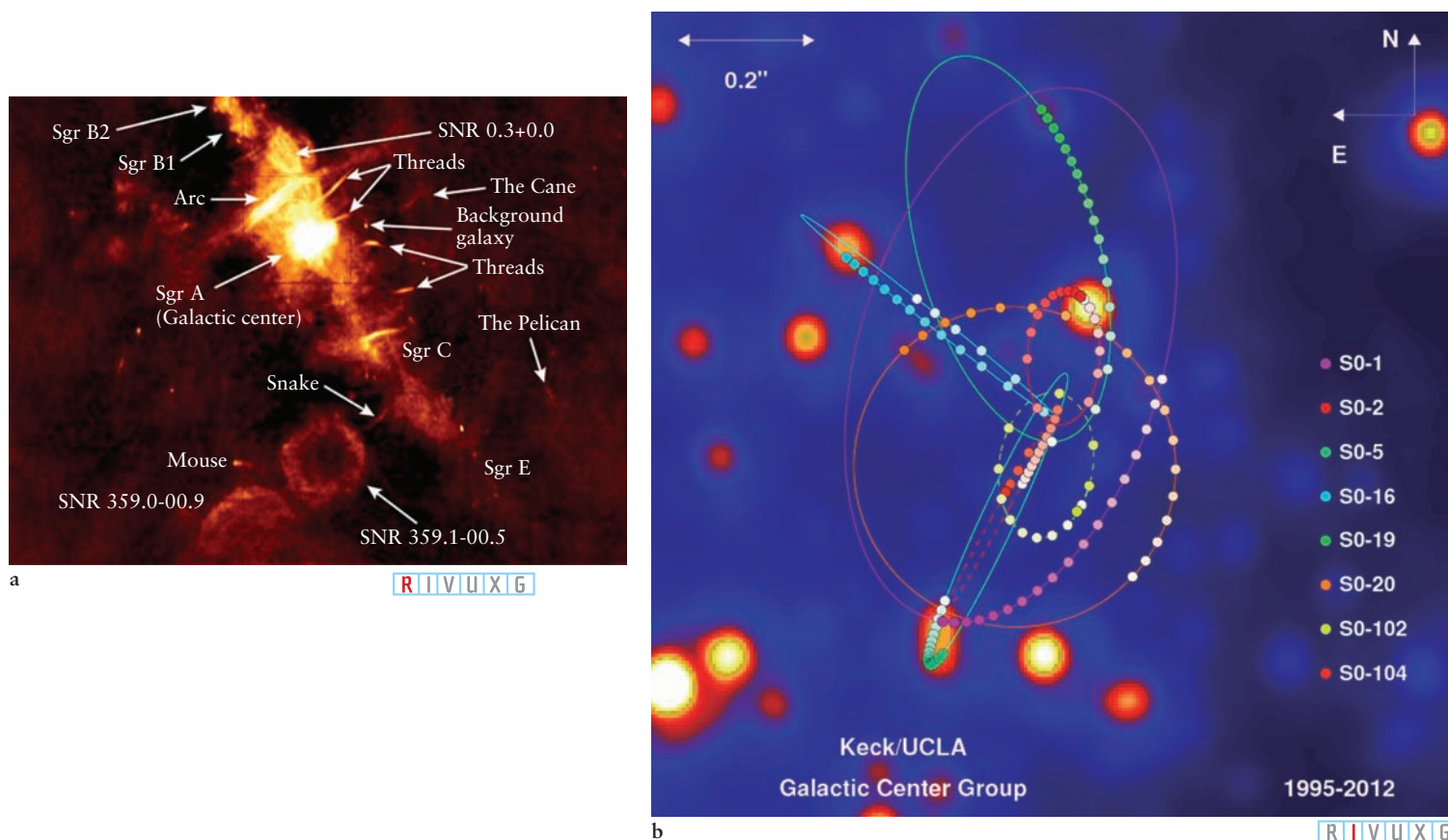


FIGURE 13-13 Two Views of the Galactic Nucleus (a) A radio image taken at the VLA of the galactic nucleus and environs. This image covers an area of the sky 8 times wider than the Moon. SNR means supernova remnant. The numbers following each SNR are its right ascension and declination. The Sgr (Sagittarius) features are radio-bright objects. (b) The colored dots superimposed on this infrared image show the motion of seven stars in the vicinity of the unseen massive object at the position of the radio source Sagittarius A*, part of Sgr A in (a). The orbits were

measured over a 17-year period. The star with the fastest known orbit goes around once every 11.5 years. This plot indicates that the stars are held in orbit by a $4.3 \times 10^6 M_{\odot}$ black hole. (a: Image processing at the Naval Research Laboratory using DoD High Performance Computing Resources. Produced by N. E. Kassim, D. S. Briggs, T. J. W. Lazio, T. N. LaRosa, J. Imamura, and S. D. Hyman. Original data are from the NRAO Very. b: These images/animations were created by Prof. Andrea Ghez and her research team at UCLA and are from data sets obtained with the W. M. Keck Telescopes.)

as it radiates away its energy (and therefore cools). In 2013, X-ray observations revealed a jet of high-energy gas particles is being emitted by Sagittarius A* perpendicular to the plane of the Milky Way. By studying this jet, astronomers have determined that the supermassive black hole is spinning in the same direction that our Galaxy is rotating.

In 2010, gamma-ray emissions in the shape of a pair of enormous bubbles were also observed centered on the galactic nucleus (Figure 13-14). They may have been created by energy emitted during a recent burst of star formation observed in that region by SOFIA (Stratospheric Observatory for Infrared Astronomy).

If the center of our Galaxy is not active and bizarre enough, recent gamma-ray observations reveal positrons either falling toward or emerging from that region at a rate of at least 10^{43} per second. (Recall that positrons have positive electric charges but are otherwise identical to electrons.) The source of these positrons is still unknown. Positrons can be detected because when a positron and an electron collide, they annihilate each other. Their mass is converted into energy as gamma rays with well-defined wavelengths. These special gamma rays have been observed emanating from the galactic center.

Infrared observations also reveal stars and gas in very rapid orbits around Sagittarius A* (Figure 13-13b). Something extremely massive must be holding this high-speed matter in such tight orbits around the galactic nucleus. Using Kepler's third law, astronomers calculate that $4.3 \times 10^6 M_{\odot}$ is needed to prevent the stars and gas from flying off into interstellar space.

The 2001 flare from Sagittarius A*, mentioned above and observed by the Chandra X-ray Observatory, rose in intensity in just a few minutes. It lasted for about 3 hours and then subsided, again in just a few minutes. The bigger that central object is, the longer the change

in brightness of the flare of X-rays will take. The duration of that flare indicated that the object at the center of our Galaxy is no wider than 1 AU across. The only known object so compact and with such a high mass is a supermassive black hole. Observations indicate that it is spinning very slowly—a Kerr black hole.

As we saw in Chapter 12, extraordinary activity is also occurring in the nuclei of most other galaxies, implying the presence of supermassive black holes at their centers as well. Astronomers are actively studying these regions in an effort to understand the complex, intriguing events that are happening in them.

Focus Question 13-3

The presence of supernova remnants at the center of our Galaxy implies what other activity is occurring in that region?

13-5 Our Galaxy's disk is surrounded by a two-shell spherical halo of stars and other matter

As mentioned at the beginning of this chapter, stars have been observed in a spherical distribution, called the **halo**, which consists of two concentric shells centered on the galactic nucleus and extending far beyond the disk (see Figure 13-10). The inner halo was originally discovered because of the globular clusters that it contains. Looking out of the plane of the Milky Way's disk and between globular clusters, we see apparently unobstructed views of distant galaxies. Most of the globular clusters orbit within about 100,000 ly of the nucleus, about twice the distance to the edge of the visible disk in which we reside.

Appearances can be deceiving. Although the brightness of the globular clusters suggests that they contain most of the stars in the halo, it turns out that about 99% of the halo's stars are isolated *halo field stars* spread all through the halo. The localized concentrations of stars in globular clusters account for only 1% of the halo's stars. It is worth noting, however, that some globular clusters are observed to contain intermediate-mass black holes, which contribute significantly to their total mass.

The globular clusters and field stars in the inner halo orbit in the same general sense that the stars in the disk orbit. The stars in the outer shell of the halo, discovered in 2007, have very little metal, implying that they are among the oldest stars in the universe. Furthermore, these latter stars orbit in the opposite sense to the motion of the disk and inner halo stars. These big differences between the two components of the halo strongly suggest that they formed from different components of the early universe. Understanding this formation process will help astronomers develop a theory of how the Milky Way developed.

The various components of the Milky Way Galaxy, including the two shells of the halo, overlap and intersect each other. For example, the disk slices through the central bulge, and globular clusters and halo field stars periodically pass through the plane of the disk.

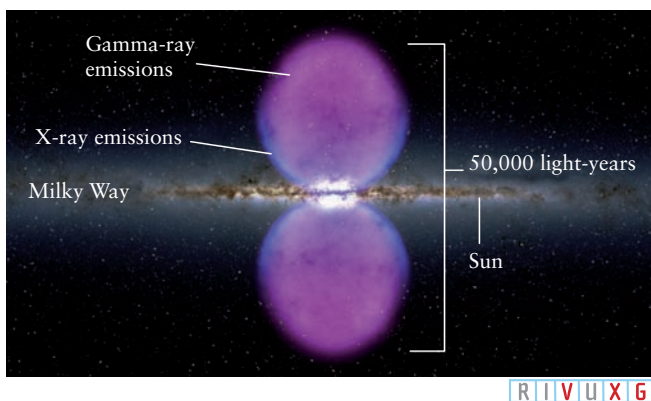


FIGURE 13-14 Gamma-Ray Bubbles in the Milky Way Shown in violet, two gamma-ray-emitting bubbles of hot gas extend perpendicularly to the plane of the Galaxy. They are centered on the galactic nucleus. X-rays are visible on the boundaries of the bubbles close to the nucleus. (NASA's Goddard Space Flight Center)

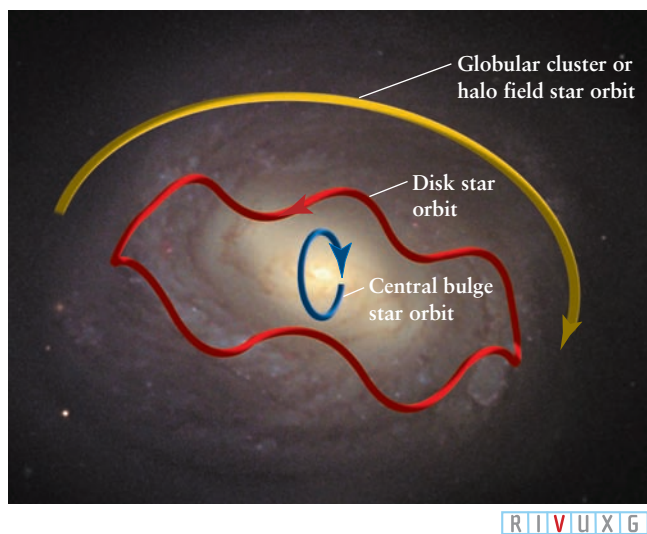


FIGURE 13-15 Orbits of Stars in Our Galaxy This disk galaxy, M58, looks very similar to what the Milky Way Galaxy would look like from far away. The colored arrows show typical orbits of stars in the central bulge (*blue*), disk (*red*), and halo (*yellow*). Interstellar clouds, clusters, and other objects in the various components have similar orbits. (NOAO/AURA/NSF)

Figure 13-15 shows the shapes of the orbits of typical central bulge, disk, and halo stars and clusters. Our solar system is presently moving at an angle of 25° from the plane of the Galaxy's disk, as illustrated by the red path in the figure. We will move above the plane about 230 ly and then return, crossing the plane every 33 million years.

Since 1994, astronomers have observed 24 small galaxies that orbit in our Galaxy's halo. The Sagittarius Dwarf and the Canis Major Dwarf (Figure 13-16) are named after the constellations in which their central regions lie. They have both spread widely through the halo and periodically pass through the disk of the Milky Way. These two galaxies are much smaller in size and mass than the Milky Way and are losing stars due to the disrupting influence of our Galaxy's strong gravitational attraction. Within the next 100 million years, they will cease to exist, their stars becoming part of the Milky Way. This process of a bigger galaxy consuming a smaller one is called **galactic cannibalism**. The most recently discovered dwarf, called the Ursa Major Dwarf spheroidal galaxy, was found in 2005. While the consumption of these smaller galaxies is believed to have begun several billion years ago, astronomers discovered in 2011 the remnants of a galaxy, called the Aquarius Star Stream, whose consumption began only 700 million years ago. Computer simulations of galactic cannibalism performed in 2010 suggest that most of the halo field stars orbiting in the Milky Way are remnants of smaller galaxies.

13-6 The Galaxy is rotating

3 Our solar system is moving at 878,000 km/h (500,000 mi/h) around the center of our Galaxy. Just as the orbital motion of the planets keeps them from falling into the Sun, the motion of the stars and interstellar clouds around the galactic center keeps these

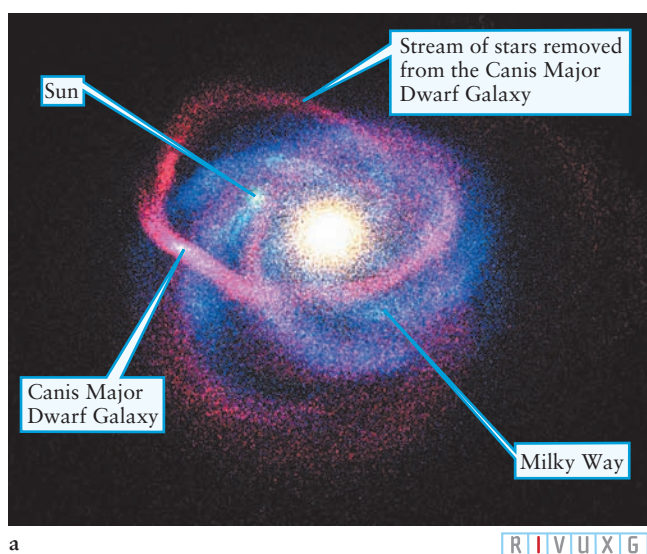
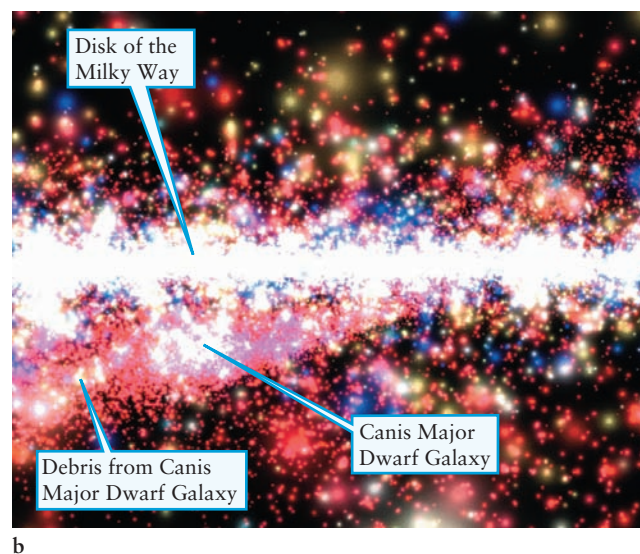


FIGURE 13-16 The Nearest Galaxy (a) The Canis Major Dwarf Galaxy is a dwarf elliptical galaxy that lies some 25,000 ly from the Milky Way. This infrared radiation-based image shows the Milky Way's spiral arms, as well as the distribution of stars being stripped from the Canis Major Dwarf Galaxy by our Galaxy's gravitational tidal force. Containing only about



1 billion stars, the Canis Major Dwarf Galaxy will be completely pulled apart within the next 100 million years or so by the Milky Way. **(b)** View from Earth of the Canis Major Dwarf Galaxy and its path of debris. (a: Nicola Martin and Rodrigo Ibata, Observatoire de Strasbourg, 2003; b: R. Ibata [Strasbourg Observatory, ULP] et al., 2MASS, NASA)

bodies apart. If the stars and clouds in our Galaxy were not in orbit, their mutual gravitational forces would have caused them to fall together and form one massive black hole billions of years ago. (Indeed, such infall of gas clouds early in the life of the universe is believed to be the origin of the supermassive black holes located at the centers of the Milky Way and other galaxies.) Just as detecting the positions of the stars and clouds has been difficult, so too has been measuring their orbital motions.

Observations show that the Sun's distance from the center of the Galaxy varies little throughout its orbit. About 80% of the stars in our neighborhood have similar-shaped orbits. The other 20% have orbits that are either taking them outward from the center or inward toward it, including at least 24 stars moving outward with enough speed to leave the Galaxy forever. The fastest of these exiting, so-called *hypervelocity stars* is moving outward with a speed of 3.2 million km/h (2.0 million mi/h).

Radio observations of 21-cm radiation from hydrogen gas provide important clues about our Galaxy's overall rotation. By measuring Doppler shifts, astronomers can determine the speed of objects toward or away from us across the Galaxy. These observations clearly indicate that our Galaxy does not rotate like a rigid body (Figure 13-17a) but rather exhibits *differential rotation*, meaning that stars at different distances from the galactic center orbit the Galaxy at different rates (Figure 13-17b).

To understand differential rotation, we focus on those stars with nearly circular orbits. Because of the stars' differential rotation in the Galaxy, the Sun is like a car on a circular freeway with the fast lane on one side and the slow lane on the other. As sketched in Figure 13-17b, stars in the fast lane (closer to the center of the Galaxy) are passing the Sun and thus appear from our vantage point to be moving in one direction, while stars in the slow lane (farther from the center of the Galaxy than our solar system) are being overtaken by the Sun and therefore appear to be moving in the opposite direction. This is like the retrograde motion of the planets discussed in Section 2-2.

Unfortunately, like the 21-cm observations, studying the motion of nearby stars and gas reveals only how fast they are moving relative to the Sun. To get a complete picture of the Galaxy's rotation, we must find out how fast the Sun itself is orbiting the center of the Galaxy. The Swedish astronomer Bertil Lindblad proposed a method of computing this speed. He noted that globular clusters do not move in the orderly pattern shown in Figure 13-17. Different globular clusters in the halo of our Galaxy orbit in different planes, and so they do not participate in the organized rotation of the objects in the Galaxy's disk. However, the

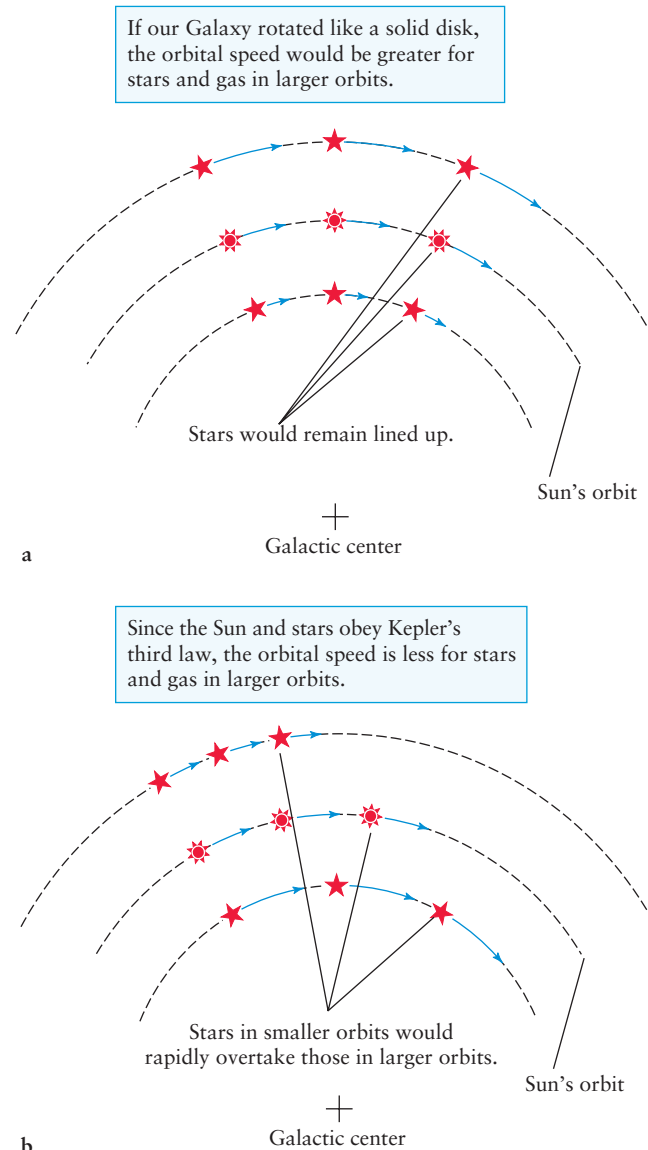


FIGURE 13-17 Differential Rotation of the Galaxy (a) If all stars in the Galaxy had the same angular speed, they would orbit in lockstep. (b) However, stars at different distances from the galactic center have different angular speeds. Stars and clouds farther from the center take longer to go around the Galaxy than do stars closer to the center. As a result, stars closer to the Galaxy's center than the Sun are overtaking the solar system, whereas stars farther from the center are lagging behind us.

combined velocities of the globular clusters around the center of the Galaxy must average to zero or else they would drift, en masse, relative to the rest of the Galaxy. Comparing the solar system's motion to the combined velocities of all the globular clusters, astronomers calculated the speed of the Sun's orbit around the galactic center given earlier.

Knowing the Sun's speed and its distance from the galactic center, astronomers can calculate the Sun's orbital period. Traveling at 878,000 km/h, our Sun takes about 230 million years to complete one trip around the Galaxy. When the solar system last passed our present location in the Milky Way, early dinosaurs of the Triassic period roamed Earth.

By combining the true speed of the Sun with the relative speed of the stars around us, as measured by radio astronomers, we can also determine the actual orbital speeds of the stars. This computation gives us the **rotation curve** of the Galaxy, a graph that shows the orbital speeds of stars and interstellar clouds at various distances from the center of the Milky Way (**Figure 13-18**).

Focus Question 13-4

The complete trip of the solar system around the Galaxy is analogous to what period in Earth's motion in the solar system?

Knowing the Sun's velocity around the Galaxy from the rotation curve, we can use Kepler's third law to estimate that the mass of the Galaxy that lies between us and the galactic nucleus is about 110 billion $M_{\odot}$.

MYSTERIES AT THE GALACTIC FRINGES

Because the band of the Milky Way extends all around Earth, we know that stars and other matter exist beyond the Sun's orbit. In recent years, astronomers have been astonished to discover just how much

matter lies farther from the center of our Galaxy than we are.

13-7 Most of the matter in the Galaxy has not yet been identified

According to Kepler's third law, out beyond the bulk of the Galaxy's visible disk, the orbital speeds of stars should decrease with distance from the Galaxy's center (see the dashed line in Figure 13-18), just as the orbital speeds of the planets decrease with increasing distance from the Sun (which contains the bulk of the solar system's mass). Observations reveal, however, that galactic orbital speeds continue to *climb* well beyond the visible edge of the galactic disk. Thus, there must be more gravitational force from the Galaxy acting on the distant stars and clouds than we can see or have taken into account.

These observations led astronomers to calculate that a surprising amount of matter must lie beyond the bulk of the visible matter (that is, the stars, gas, and dust) in the Galaxy. Nearly 90% of the mass of our Galaxy has yet to be located. If we include the mass of the as-yet-unidentified matter, the Milky Way's total mass must be around $1.2 \times 10^{12} M_{\odot}$.

To make matters even more mysterious, much of this mass is in the form of dark matter: Whatever it is, very little of it shows up on images at any wavelength. Given its gravitational effect on the stars and gas in the Galaxy that we can detect, astronomers deduce that this dark matter is spherically distributed all around the Galaxy in a massive halo extending at least 12 times farther from the galactic nucleus than we are, far beyond the halo defined by the visible globular clusters and halo field stars (see inset in Figure 13-10). Small fractions of the dark matter are composed of neutrinos and white dwarfs, but the nature of the rest of it—whether black holes, gas, Jupiterlike bodies, brown dwarfs, or something far more exotic—has yet to be determined. The unidentified matter in our Galaxy is sometimes called the *missing mass*. This is a poor name because the matter is not missing; we just do not yet know its nature.

Some searches for dark matter rely on the principle from general relativity that matter causes passing light to change direction (see Figure 12-32). The Galaxy's dark matter may contain relatively massive objects, like brown dwarfs or black holes. If so, these objects will cause the light from more distant stars to change direction as the dark objects pass between those stars and Earth (**Figure 13-19a**). Such objects are called *MACHOs* (massive compact halo objects). Indeed, as a dark object moves directly between a more distant star and Earth, the star's light is focused toward us, causing the star to brighten. Called **microlensing**, this brightening of distant stars for several weeks has been observed (Figure 13-19b). Microlensing observations indicate that only a

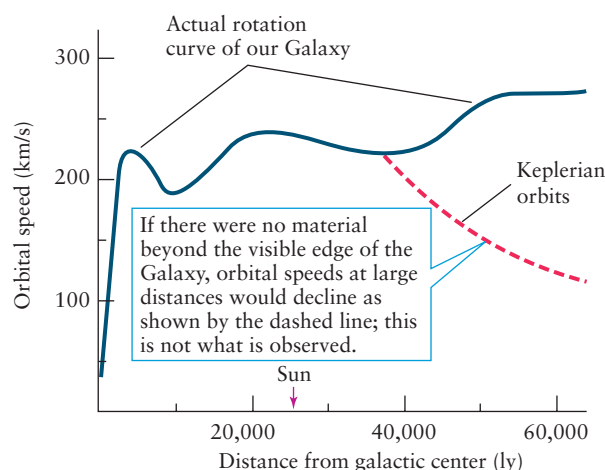
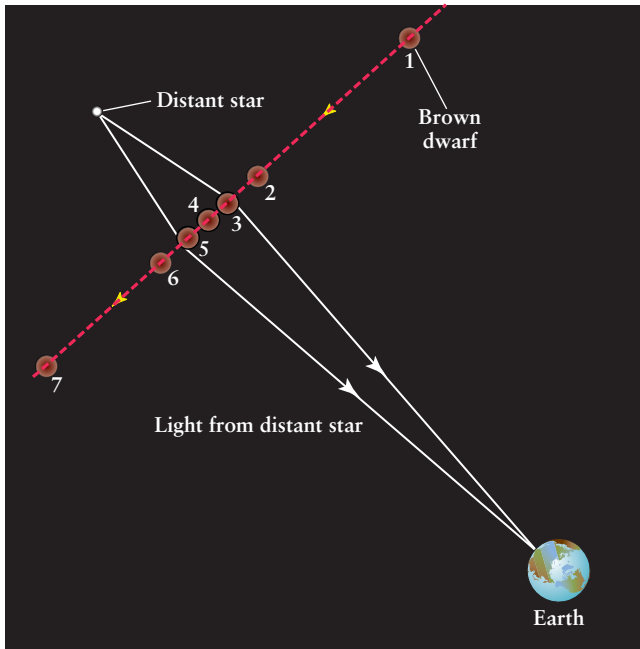
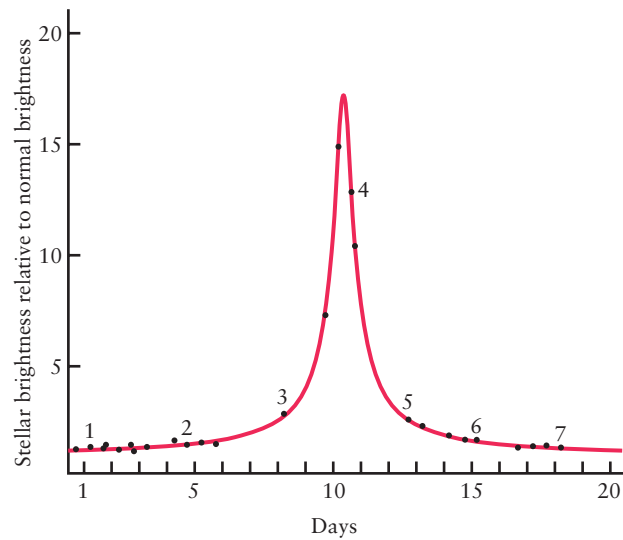


FIGURE 13-18 The Galaxy's Rotation Curve The blue curve shows the orbital speeds of stars and gas in the Galaxy, and the dashed red curve shows Keplerian orbits that would be caused by the gravitational force from all the known objects in the Galaxy. Because the data (blue curve) do not show any decline, there is, apparently, an abundance of dark matter that extends to great distances from the galactic center. This additional mass gives the outer stars higher speeds than they would have otherwise.



a

FIGURE 13-19 Microlensing by Dark Matter in the Galactic Halo (a) Gravitational fields cause light to change direction. A white dwarf, brown dwarf, or black hole in the Galaxy's halo passing between Earth and a more distant star will focus



b

the starlight in our direction, making distant objects appear brighter than they are normally. (b) The light curve of the gravitational microlensing of light from a star in the Galaxy's central bulge by an intervening object.

small fraction of the Galaxy's dark matter is in the form of these MACHOs.

TYPES OF GALAXIES

Despite the incredible number of galaxies, there is surprising consistency in their overall shapes. Edwin Hubble began cataloging their appearance in the 1920s. The **Hubble classification** of galaxies—spirals, barred spirals, lenticulars, ellipticals, irregulars, and their subclasses—is still used today.

13-8 The winding of a spiral galaxy's arms is correlated to the size of its central bulge

Spiral galaxies without bars (often called *normal spiral galaxies*) are characterized by a central bulge and arched lanes of stars and interstellar clouds, which appear as arms that spiral out from the bulge. Observations of many spiral galaxies reveal that the more tightly wound the spiral, the larger the central bulge.

As shown in **Figure 13-20**, spirals with tightly wound spiral arms (and fat central bulges) are called



a Sa (NGC 1357)

R I V U X G



b Sb (M81)

R I V U X G



c Sc (M100)

R I V U X G

FIGURE 13-20 Spiral Galaxies (Nearly Face-on Views) Edwin Hubble classified spiral galaxies according to the tightness of their spiral arms and the (correlated) sizes of their central bulges. Sa galaxies have the largest central bulges and the most tightly wound spiral arms, whereas Sc galaxies have the smallest

central bulges and the least tightly wound arms. (a: Adam Block/Steve Mandel/Jim Rada and Students/NOAO/AURA/NSF; b: NASA, ESA, and the Hubble Heritage Team [STScI/AURA], acknowledgment: A. Zezas and J. Huchra [Harvard-Smithsonian Center for Astrophysics]; c: Australian Astronomical Observatory/David Malin Images)

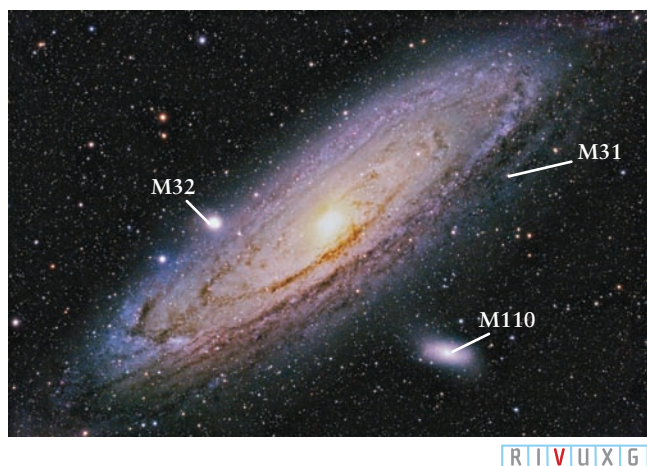


FIGURE 13-21 Andromeda (M31) Andromeda is a beautiful spiral galaxy and the only galaxy easily visible to the naked eye from Earth's northern hemisphere. It has dim, red giant stars (not visible here) extending half a million light-years from its nucleus. Without a telescope, Andromeda appears to be a fuzzy blob in the constellation of the same name. Located only 2.5 Mly (0.77 Mpc) from us, Andromeda is gravitationally bound to the Milky Way, and it covers an area in the sky roughly 5 times as large as the full Moon. Two other galaxies, M32 and M110, are also labeled on this photograph. The points of light that pepper the image are stars in our Galaxy. (Astrophotography by Terry Hancock. Used with permission of Getty Images.)

Sa (for spiral type a) *galaxies*. Those with moderately wound spiral arms (and a moderate central bulge) are *Sb galaxies*. Loosely wound spirals (with small central bulges) are *Sc galaxies*. A typical spiral galaxy contains about 100 billion stars and measures nearly 10^5 light-years in diameter.

The closest spiral galaxy to the Milky Way is the Andromeda Galaxy, M31 (Figure 13-21). M31 is visible to the naked eye as a fuzzy blob in the constellation of Andromeda. Rather than being face-on or edge-on, the disk of this galaxy is tilted as seen from Earth, typical of many spiral galaxies. Some are tilted so much that their spiral arms are not evident. We can still classify these galaxies as specific spiral types by observing the sizes of their central bulges. For example, M104 (Figure 13-22a) has a huge central bulge. It must therefore be an *Sa* galaxy with tightly wound arms. An *Sb* galaxy (Figure 13-22b) has a smaller central bulge. The

tiny central bulge of an *Sc* galaxy (Figure 13-22c) is hardly noticeable at all in an edge-on view.

Besides the degree of winding, the overall appearance of individual spiral arms typically has one of two structures. In some galaxies, called *flocculent spirals* (from the word meaning “fleece”), the spiral arms are broad, fuzzy, chaotic, and poorly defined (Figure 13-23a). Other galaxies, called *grand-design spirals*, exhibit beautiful arching arms outlined by brilliant H II regions and OB associations. In these galaxies, the spiral arms are thin, delicate, graceful, and well defined (Figure 13-23b).

At first glance, it may seem that the shapes of spiral galaxies are created by the orbital motions of strings of stars, with the stars closer to the center of the galaxy orbiting more rapidly and thereby leading the outer edges of the spirals. But, in this case, common sense leads us astray. Although the stars closer to the center of the galaxy do orbit faster than those farther away, spiral



a M104: an *Sa* galaxy



b NGC 891: an *Sb* galaxy



c NGC 5907: an *Sc* galaxy

FIGURE 13-22 Spiral Galaxies Seen Nearly Edge-on from the Milky Way (a) Because of its large central bulge, this galaxy (called the Sombrero Galaxy) is classified as an *Sa*. If we could see it face-on, the spiral arms would be tightly wound around a voluminous bulge. (b) Note the

smaller central bulge in this *Sb* galaxy. (c) At visible wavelengths, interstellar dust obscures the relatively insignificant central bulge of this *Sc* galaxy. (a: NASA and the Hubble Heritage Team [STScI/AURA]; b: Malin/IAC/RGO; c: Bran Ehrhorn/Adam Block/NOAO/AURA/NSF)



a M33: a spiral galaxy with flocculent spiral arms

R I V U X G



b M74: a grand-design spiral galaxy

R I V U X G

FIGURE 13-23 Variety in Spiral Arms The differences in spiral galaxies suggest that at least two mechanisms create spiral arms. **(a)** This flocculent spiral galaxy has fuzzy, poorly defined

spiral arms. **(b)** This grand-design spiral galaxy has well-defined spiral arms. (a: NASA; b: Gemini Observatory/AURA)

arms cannot be orbiting strings of stars. If they were, the spiral arms should “wind up” from the inside out (**Figure 13-24**), wrapping themselves tightly around the nucleus. While this action would take too long for us to observe in a single galaxy, hundreds of thousands of spiral galaxies have been observed, and all are seen in just a few states of “winding” (see Figure 13-20), which is not consistent with this model.

Furthermore, if star motions caused spirals, then after a few galactic rotations, the spiral structure should disappear altogether. Why, then, do we observe spiral arms in our Galaxy and many others? This *winding dilemma* led astronomers to two explanations, one for flocculent spirals and the other for grand-design spirals.

Focus Question 13-5

Comparing it to the galaxies in Figure 13-20 or 13-22, what is the spiral classification of M74, shown in Figure 13-23b?

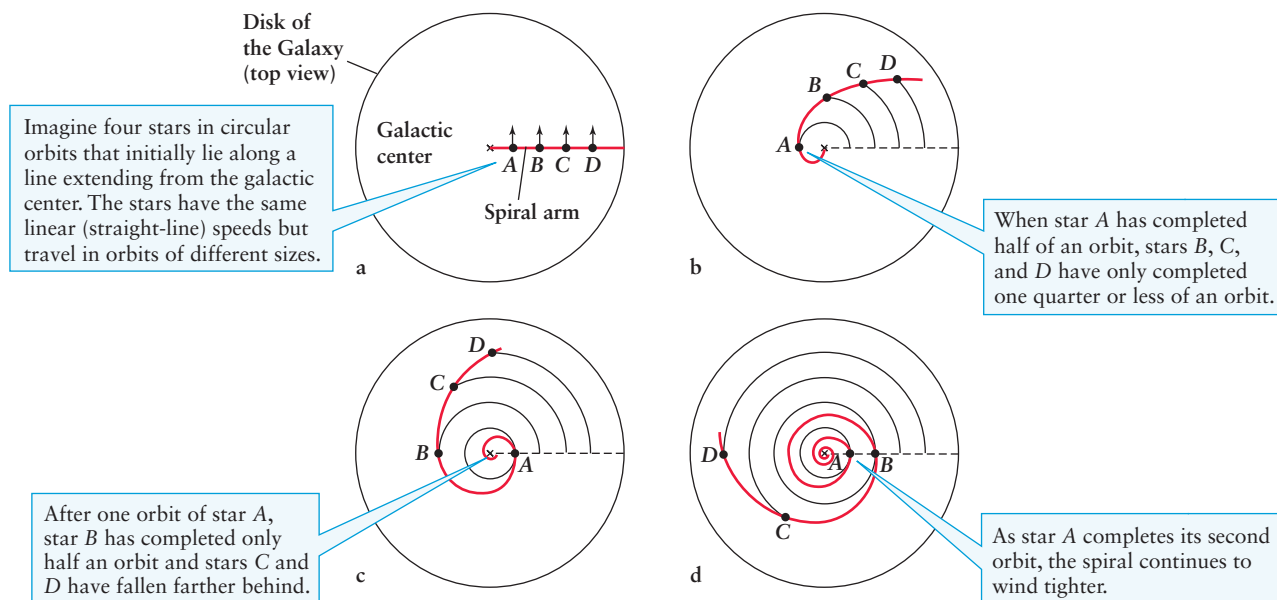


FIGURE 13-24 The Winding Dilemma The rotation curve of the disk stars in our Galaxy indicates that most of them have the same linear (straight-line) speed. Those stars farther from the center take longer to go around because they have a greater distance to travel at the same speed than stars closer to the center of the Galaxy, which orbit in smaller circles. All four

dots in these drawings have circular orbits at the same linear speed. Think of these dots as **(a)** a few stars initially in a straight line. As time goes on, the outer stars are left behind, creating **(b–d)** a spiral shape that becomes more and more tightly wound. Such tightening is not observed in our Galaxy or in other galaxies.

13-9 Explosions create flocculent spirals, and waves create grand-design spirals

Flocculent spiral galaxies Consider a disk galaxy so young that it has not yet formed spiral arms. Stars begin forming in one of its giant molecular clouds. The radiation and winds from these stars compress the nearby interstellar gas, triggering the formation of additional stars. Moreover, massive stars quickly explode as supernovae and produce shock waves, which further compresses surrounding gases until it is Jeans unstable (see Section 4-2). This gas collapses and forms new stars. Shock waves are very powerful compressions of the gas, like the sonic booms created by lightning, the snap of a whip, or a supersonic aircraft. Thus, as new stars trigger the birth of still other stars, the star-forming region grows, a process called *self-propagating star formation*.

The continuing birth of stars accounts for spiral arms in flocculent spiral galaxies. The galaxy's differential rotation (inner regions orbiting faster than outer regions; see Figure 13-24) drags the inner edge of each young region ahead of its outer edge, spreading the star-forming region into a spiral arm. This region of the galaxy is bright because it is highlighted by its brilliant O and B stars and by the nebulae that these stars cause to glow. However, high-mass O and B stars explode relatively quickly, and each spiral arm dims before it winds up (that is, there is no

winding dilemma here). As old arms fade, new star-forming regions appear and create new spiral arms.

Where self-propagating star formation is the only process creating spiral arms in a galaxy, bits and pieces of these arms appear constantly, only to disappear as the bright, massive stars in them die. This process creates chaotic spiral arms, such as those observed in flocculent galaxies (see Figure 13-23a) but not the smooth spirals of grand-design galaxies. Thus, self-propagating star formation cannot be the whole story.

Waves create grand-design spiral galaxies Our search for the mechanism that generates grand-design spirals begins in a pond. If you throw a rock into the water, you create ripples. As you know from experience, the ripples move outward in concentric rings from where the rock struck (Figure 13-25a). But suppose the pond water is rotating when you throw in a rock. Now what shape do the ripples have? According to calculations done by astronomer Bertil Lindblad in the 1920s, ripples created in a rotating disk-shaped system of liquid (the pond) or gas and dust (a disk galaxy) will be spiral. These ripples are called **spiral density waves**.

Russian scientists in the 1970s tested Lindblad's model. Lacking the computer power to create simulations of galaxies, they threw pebbles into pie pans of water rotating on phonograph turntables. The resulting water wave patterns were indeed spirals, as drawn in Figure 13-25b. Furthermore, spiral density waves never

Focus Question 13-6

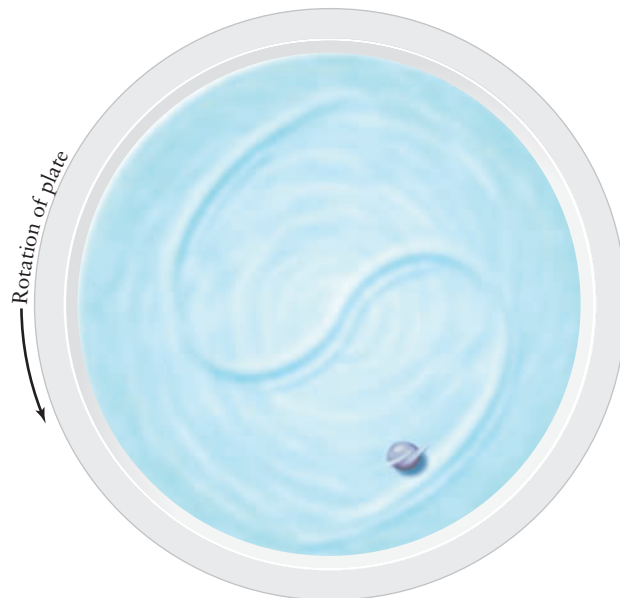
Fill in the blanks: If analogous events occurred on Earth, we would call self-propagating star formation a _____ reaction or a _____ effect.



a

RITVUXIG

FIGURE 13-25 Ripples in Water (a) The usual circular ripples expanding from the place where a rock was thrown into the



b

water. (b) Ripples in rotating water create spiral arms, as do ripples in the gas and dust of a disk galaxy. (a: Corbis)

wind up. In the case of the gas and dust in a galaxy, these waves orbit in a rigid spiral pattern at about half the average speed of the interstellar medium through which they move.

In the mid-1960s, two American astrophysicists looked carefully at how galactic spiral density waves traveling through the disk of a galaxy lead to spiral arms. C. C. Lin and Frank Shu argued that rigidly rotating spiral-shaped density waves cause interstellar gas and dust passing through them to temporarily pile up. This piling-up happens because, unlike water waves, which move up and down as they travel along, spiral density waves in the gas and dust of a galaxy are compressional waves, just like sound. As a result, they cause interstellar gas and dust to become compressed and, therefore, Jeans unstable as this material passes through the waves. The Jeans instabilities (see Section 4-2) lead to new star formation in the spiral density waves, creating spiral arms.

Lin and Shu demonstrated that a spiral arm is a galactic traffic jam. Imagine a slow-moving truck on a busy freeway. The cars normally cruise at 100 km/h (about 65 mi/h), but the truck causes a bottleneck. The cars slow down temporarily to avoid hitting other cars and the slowly moving truck. As seen from the air, there is a noticeable congestion of cars around the truck (Figure 13-26). An individual car spends only a few moments in the moving traffic jam before resuming its usual speed, but the traffic jam itself lasts all day long.

The vehicles in the traffic jam represent interstellar gas and dust entering a spiral arm. The truck represents a small part of a spiral density wave. Like the cars catching up to and passing the truck, the metal-rich interstellar medium sweeps through the more slowly moving spiral density waves. The waves compress some of this interstellar gas and dust, which contract to form new, metal-rich stars. We see spiral arms because they contain bright O and B stars and copious quantities of dust and gas that these stars illuminate. The sprawling dust lanes in Figure 13-23b attest to the passage of that material through spiral density waves. While all of this occurs, the density wave maintains its shape and, overall, the spiral arms in these galaxies are better defined than in the flocculent galaxies discussed earlier.

4 A few high-mass stars form in most open clusters, such as those created in spiral density waves. So why are spiral galaxies not eventually filled with enough of these bright stars to wash out the spiral arms and to make the disk uniformly bright? The answer is that the massive stars highlighting each density wave are short-lived and they explode before they finish passing through it. The remaining, longer-lived, lower-mass stars, like our Sun, fill the space between the spiral arms without emitting

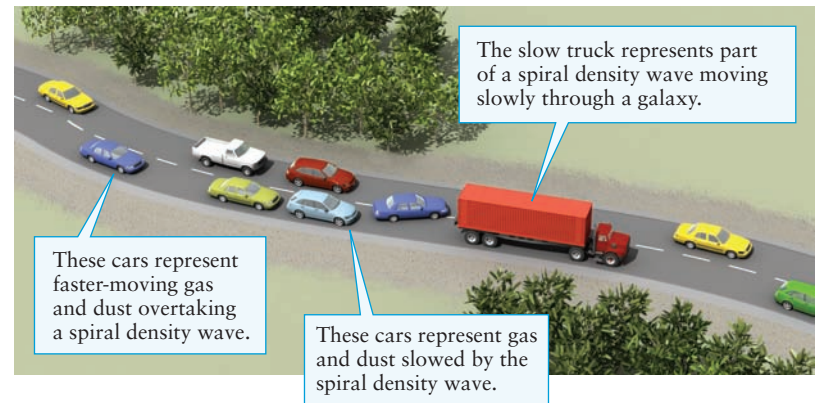


FIGURE 13-26 Compression Wave in Traffic Flow When normal traffic flow is slowed down, cars bunch together. In a grand-design galaxy, a density wave moves through the stars and gas. The wave is merely a region of slightly denser matter, which, in turn, creates more gravitational force. This force compresses the gas passing through it, thereby causing star formation, which highlights the spiral density wave.

as much light (see Figure 13-23b). As pronounced as spiral arms may appear, they contain only 5% more stars than are found between them.

It takes an enormous amount of energy from the spiral density waves to compress interstellar gas and dust. After a billion years or so, even spiral density waves would begin to fade away. Some driving mechanisms must keep them going, akin to throwing more rocks in the pond to maintain its ripples. We are not yet completely certain what those mechanisms are, but one likely explanation for reinvigorating the galactic spiral structure is the passage of a nearby companion galaxy. As this nearby galaxy periodically passes close by, its gravitational attraction pulls on the gas, stars, and dust of the spiral galaxy and causes this latter galaxy to generate new density waves. Indeed, grand-design galaxies are usually found in the presence of a companion galaxy.

Although most spiral galaxies have two arms, a sizable minority has more. Astronomers have not yet established why the numbers of arms vary.

Because new stars are formed from interstellar gas and dust, it is plausible that galaxies with different amounts of this material will have different star formation rates and, therefore, different overall structures. Infrared and radio observations reveal that Sa galaxies typically contain about 4% gas and dust, whereas Sb galaxies contain 8%, and Sc galaxies contain 25%. Continually improving computer simulations of galaxies reveal more and more of the details of how various concentrations of gas and dust affect galactic structure. Figure 13-27 summarizes the activity in a spiral galaxy.



FIGURE 13-27 Dynamics of a Grand-Design Spiral Galaxy
This figure summarizes the activities taking place in a grand-

design spiral galaxy. (NASA and ESA; inset: S. Kafka and K. Honeycutt, Indiana University/WIYN/NOAO/NSF)

13-10 Bars of stars run through the central bulges of barred spiral galaxies, and some disk galaxies, the lenticulars, lack spiral arms

The Milky Way is a **barred spiral galaxy**, a spiral galaxy with a bar of stars and gas crossing through the central bulge. Bars form in some disk galaxies when stars near the centers of these galaxies that were originally following circular paths are pulled into more elliptical orbits. When

enough stars are in highly elliptical orbits, they create the bars in barred spiral galaxies. These stars then gravitationally attract gas into the same orbits. Some of this gas is channeled toward the supermassive black holes residing in the centers of most, if not all, spiral galaxies.

Observations reveal that about 70% of nearby spiral galaxies have bars, while only 20% of more distant spiral galaxies do. The farther away that we see things, the longer the light from them has been traveling toward us. Since light travels at a finite speed, this means that



a M58: an SBa galaxy

R I V U X G



b M83: an SBb galaxy

R I V U X G



c NGC 1365: an SBc galaxy

R I V U X G

FIGURE 13-28 Barred Spiral Galaxies As with spiral galaxies, Edwin Hubble classified barred spirals according to the tightness of their spiral arms (which correlates with the sizes of their central bulges). (a) SBa galaxies have the most tightly wound spirals and largest central bulges, (b) SBb galaxies have moderately tight spirals and medium-sized central

bulges, and (c) SBc galaxies have the least tightly wound spirals and the smallest central bulges. (a: Johan H. Knapen and Nk Szymanek, University of Hertfordshire; b: European Southern Observatory; c: Jean-Charles Cuillandre/Canada-France-Hawaii Telescope/Science Source)

the farther objects are from Earth, the farther back in time we are seeing them. The fact that fewer of the more distant galaxies have bars indicates that the number of barred spiral galaxies is increasing.

Taking all observed spirals into account, about one-third of them are barred spirals, the remainder being the normal spirals. The two main arms in barred spirals typically extend from the ends of the bar rather than from the central bulge itself. The largest known spiral galaxy is a barred spiral, NGC 6872, which is 522,000 ly across, more than 5 times the size of the Milky Way.

Edwin Hubble found that the winding of the spiral arms in barred spirals also correlates with the size of the central bulge (Figure 13-28), just as for normal spirals. An *SBa* (for spiral, barred, type a) galaxy has a large central bulge (and tightly wound spiral arms). Likewise, a barred spiral with moderately wound spiral arms (and a moderate central bulge) is an *SBb* galaxy, and an *SBc* galaxy has loosely wound spiral arms (and a tiny central bulge). The Milky Way is classified as an *SBab*, which indicates that the winding of its arms is between that of an *SBa* and an *SBb*. *Sa* and *SBa* are sometimes called *early-type galaxies*, while *Sc* and *SBc* are called *late-type galaxies*.

The motions of the stars in many nearby galaxies have been measured from Doppler shifts of their spectra. In all spiral and barred spiral galaxies discovered to date, except one, the arms trail around behind as the galaxies rotate. These galaxies are thus called **trailing-arm spiral galaxies**. The exception is the galaxy NGC 4622. This galaxy has the points of its spirals leading the motion of the arms. It is a *leading-arm spiral*. Calculations predict that the leading arms are unstable and that this galaxy will eventually switch and become a trailing-arm spiral.

A few disk galaxies, both unbarred and barred, lack spiral arms. These are classified as **lenticular galaxies**,

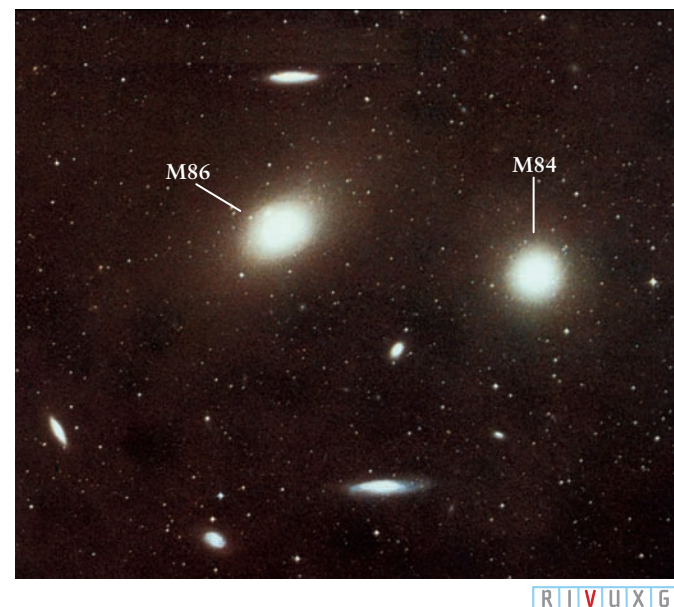
because they look like lenses. Unbarred lenticulars are labeled *S0*, while barred lenticulars are denoted *SB0*.

13-11 Elliptical galaxies display a wide variety of sizes and masses

5 Elliptical galaxies, named for their distinctive shapes, have no spiral arms. They range tremendously in size and mass—from the biggest to the smallest galaxies in the universe. Figure 13-29 includes two

Focus Question 13-7

Why do you think that spiral arms do not extend farther inward than the bar in barred spiral galaxies?



R I V U X G

FIGURE 13-29 Giant Elliptical Galaxies The Virgo cluster is a rich, sprawling collection of more than 2000 galaxies about 50 million light-years from Earth. Only the center of this huge cluster appears in this photograph. The two largest galaxies in the cluster are the giant elliptical galaxies M84 and M86. (Australian Astronomical Observatory/David Malin Images)

giant elliptical galaxies, M84 and M86, that form part of a cluster of galaxies in the constellation Virgo. These enormous giant elliptical galaxies are each about 2 million light-years in diameter, 20 times the diameter of the Milky Way Galaxy.

Giant ellipticals, containing some 10 trillion solar masses, are rare compared to other types of galaxies, whereas *dwarf elliptical galaxies* are extremely common. Dwarf ellipticals are only a fraction of the size of an average elliptical galaxy and contain so few stars—only a few million—that we see these galaxies as nearly transparent. Because you can actually see straight through the center of a dwarf elliptical galaxy and out the other side (Figure 13-30), distant ones are hard to detect. In 1999 astronomers discovered the existence of dwarf galaxies that are much more compact and have stars orbiting much more closely together than traditional dwarf elliptical galaxies. Whereas traditional dwarf elliptical galaxies are a few thousand light-years in diameter, *ultracompact dwarfs* are a few hundred light-years across. Ultracompact dwarfs represent groups of stars intermediate in size and number between globular clusters (see Section 11-14) and traditional dwarf galaxies. Many more dwarf and ultradwarf ellipticals undoubtedly exist than have been identified, and the uncertainty in their numbers is a major reason that we do not know the total number of galaxies in the visible universe.

Hubble subdivided elliptical galaxies according to how round or oval they look. The roundest elliptical galaxies are called *E0 galaxies*, whereas the most elongated are *E7 galaxies*. These latter are about 3 times longer than they are wide. Elliptical galaxies of intermediate elongation are numbered E1 to E6 (Figure 13-31).

The Hubble scheme classifies galaxies solely by their appearance from Earth. But whenever we observe anything in the sky, we are seeing a two-dimensional view of a three-dimensional object. In the case of elliptical galaxies, we observe length and width, but we have no way of knowing anything about the third dimension of

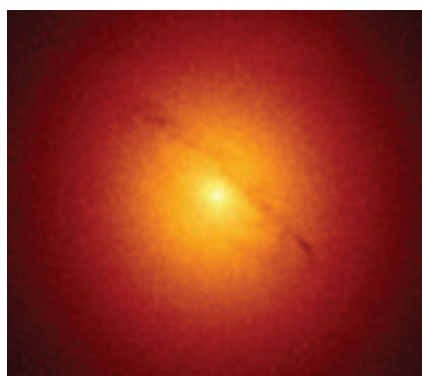


RITVUXG

FIGURE 13-30 A Dwarf Elliptical Galaxy This nearby E4 galaxy, called Leo I, is about 600,000 light-years from Earth. It is only 3000 light-years in diameter and is so sparsely populated with stars that you can see right through its center. Leo I is a satellite galaxy of the Milky Way. (Australian Astronomical Observatory/David Malin Images)

depth. What looks like an E0 galaxy (basically circular) might actually be egg-shaped when viewed from another angle. Conversely, an elongated E7 galaxy might look circular when viewed down its long axis.

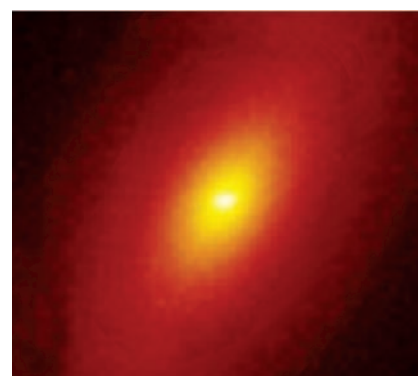
Elliptical galaxies look far less dramatic than their spiral and barred spiral cousins because they contain relatively little interstellar gas and dust. For example, the only gas visible in M105 (Figure 13-31a) is the dark line running from upper left to lower right. All the orange images are stars. Because stars form in interstellar clouds, relatively few stars should be forming in ellipticals compared to the numbers forming in spirals. Observations of spectra confirm the hypothesis that elliptical galaxies contain primarily Population II, intermediate-mass, long-lived stars. However, billions of years ago,



a E0 (M105)



b E3 (NGC 4406)



c E6 (NGC 3377)

RITVUXG

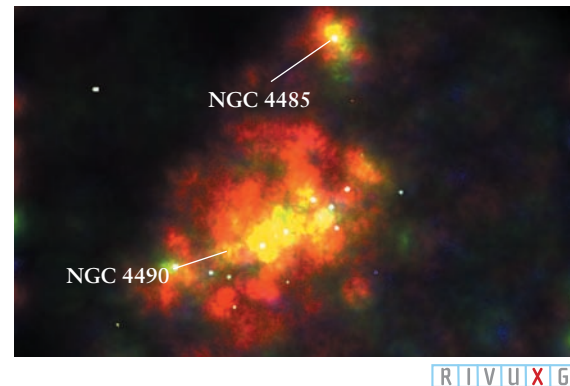
FIGURE 13-31 Elliptical Galaxies Hubble classified elliptical galaxies according to how round or elongated they appear. An E0 galaxy is round; a very elongated elliptical galaxy is an E7.

Three examples are shown. (a and c: Karl Gebhardt [University of Michigan], Tod Lauer [NOAO], NASA; b: Jean-Charles Cuillandre/Canada-France-Hawaii Telescope)



a Large Magellanic Cloud, an Irr I galaxy

FIGURE 13-32 Irregular Galaxies (a) At a distance of only 179,000 light-years, the Large Magellanic Cloud (LMC), an Irr I irregular galaxy, is passing close to our Milky Way Galaxy. About 62,000 light-years across, the LMC spans 22° across the sky, about 44 times the angular size of the full Moon. Note the huge H II region (called the Tarantula Nebula or 30 Doradus). Its diameter of 800 light-years and mass of 5 million solar masses



b NGC 4485 (Irr II) and NGC 4490 (Sc) galaxies

make it the largest known H II region. (b) The small irregular (Irr II) galaxy NGC 4485 interacts with the highly distorted Sc galaxy NGC 4490, also called the Cocoon Galaxy as shown in this X-ray image. This pair is located in the constellation Canes Venatici. (a: Australian Astronomical Observatory/David Malin Images; b: NASA/MSFC)

many ellipticals contained bright, massive stars that have since exploded as supernovae. We know this because the Chandra X-ray Observatory has discovered large numbers of remnant black holes and neutron stars in several of these galaxies.

13-12 Galaxies without global structure are called irregular

Edwin Hubble found some galaxies that cannot be classified as spirals, barred spirals, lenticulars, or ellipticals. He called these **irregular galaxies**. They are generally rich in interstellar gas, dust, and both young and old stars. Irregular galaxies with numerous OB associations and only hints of organized structure are denoted *Irr I*. The nearby Large Magellanic Cloud (LMC) (Figure 13-32a) and the Small Magellanic Cloud (SMC) are examples of Irr I galaxies. (Although they are among the closest galaxies to us, neither of them orbits the Milky Way.) Both can be seen with the naked eye from southern latitudes.

Occasionally, irregulars are observed that appear highly distorted and completely asymmetrical, as though created by collisions between galaxies or by violent activity in their nuclei. These galaxies are denoted *Irr II*, as shown by NGC 4485 (Figure 13-32b). Irregular galaxies are typically smaller and less massive than spirals, containing between about 10^8 and $3 \times 10^{10} M_\odot$.

13-13 Hubble presented spiral and elliptical galaxies in a tuning fork-shaped diagram

Edwin Hubble connected the four regularly shaped types of galaxies—spirals, barred spirals, lenticulars, and el-

lipticals—in a diagram shaped like a tuning fork (Figure 13-33). For want of any better scheme, the irregular galaxies are sometimes placed between the ends of the tuning fork prongs of the Hubble diagram.

Although Hubble did not intend it when he designed his tuning fork diagram, the idea that one type of galaxy may change into another type

Focus Question 13-8

Instead of being classified an E0, what other classification could M105 in Figure 13-31a possibly have?

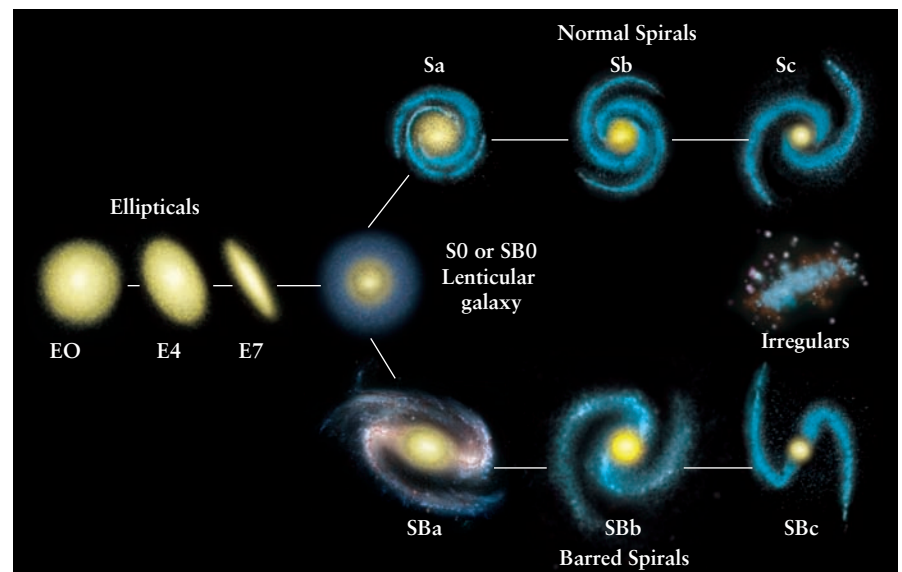


FIGURE 13-33 Hubble's Tuning Fork Diagram Hubble summarized his classification scheme for galaxies with this tuning fork diagram. Elliptical galaxies are classified by how oval they appear, whereas spirals and barred spirals are classified by the sizes of their central bulges and the correlated winding of their spiral arms. An S0 or SB0 galaxy, also called a lenticular galaxy, is an intermediate type between ellipticals and spirals. It has a disk but no spiral arms.

Focus Question 13-9
The Large Magellanic Cloud is visible to the naked eye in the southern hemisphere. From Figure 13-32a, for which terrestrial objects do you think it could be mistaken?

has been in and out of favor with astronomers for decades. As we will see later in this chapter and in Chapter 14, extensive observations by the Hubble Space Telescope and by ground-based telescopes reveal that interactions between galaxies sometimes do lead to changes in their structures. For example, when two disk galaxies merge, they often morph into a giant elliptical galaxy. Since their formation, about three-quarters of all galaxies have maintained their structures, whereas the structures of the remaining galaxies have changed (for example, from spiral to elliptical). The properties of the different types of galaxies are summarized in **Table 13-1**.

In Chapter 14, we will explore the observational evidence for the formation and early evolution of galaxies. Nevertheless, it is worth summarizing here what astronomers know about why the galaxies have the structures we observe. With the exception of dwarf ellipticals, galaxies formed from smaller ensembles of stars, gas, and dust. The underlying dark matter that clumped together in the young universe drew this matter together, often assisted by supermassive black holes that formed early in the life of the universe. These black holes became the nuclei of the galaxies. The more massive a black hole is, the larger a disk galaxy's central bulge. Recent observations show that it often took several billion years for disk galaxies to accumulate most of their gas and settle down into the stable star-forming systems we see today.

Focus Question 13-10
What physical property of matter discussed in Chapter 2, besides gravitation, is essential in forming all spiral galaxies?

Normal (unbarred) spiral galaxies formed when swirling eddies of gas-rich matter collided and formed a disk centered on an especially massive black hole and lots of dark matter. When the amount of dark matter and the mass

of the central black hole were too low, the matter in the central regions of the disk had bar-shaped orbits, creating the early barred spiral galaxies.

When the collision of galaxy-forming gases led to especially rapid star formation, the resulting galaxies were elliptical or irregular. Giant elliptical galaxies were (and still are) formed from the collisions of two spiral galaxies (of any type).

Galaxy collisions do not just occur in pairs. Stephan's quintet, discovered in 1877 by the French astronomer Edouard Stephan, is a group of five galaxies bound together. Four of its members are undergoing multiple collisions with each other. They will eventually merge to form a single giant elliptical galaxy. This is not a unique interaction: In 2002, the Hubble Space Telescope observed four galaxies colliding simultaneously.

CLUSTERS AND SUPERCLUSTERS

As we consider the vastness of galaxies, it is hard to imagine that structures this huge orbit each other in groups. But they do.

13-14 Galaxies exist in clusters that may form still larger clumps called superclusters

6 Galaxies are not scattered randomly throughout the universe, but rather they are grouped together in **clusters**. The Hercules cluster (**Figure 13-34**), so called because it is located in the constellation of Hercules, is typical. Observations of galactic motion indicate that members of a cluster of galaxies are gravitationally bound together: The galaxies in a cluster orbit each other and, occasionally, even collide. Calculations based on observations of the Andromeda Galaxy's velocity strongly suggest that it and our Milky Way will collide in 4 billion years.

Clusters of galaxies are themselves grouped together in huge associations called **superclusters**. A typical su-

TABLE 13-1 SOME PROPERTIES OF GALAXIES

	Spiral (S) and barred spiral (SB) galaxies	Elliptical galaxies (E)	Irregular galaxies (Irr)
Mass ($M_{\odot}$)	10^9 to 4×10^{11}	10^7 to 10^{13}	10^8 to 3×10^{10}
Luminosity ($L_{\odot}$)	10^8 to 2×10^{10}	3×10^5 to 10^{11}	10^7 to 10^9
Diameter (ly)	1.6×10^4 to 8×10^5	3×10^3 to 6.5×10^5	3×10^3 to 3×10^4
Stellar populations	Disk: young Population I central bulge; halo: Population II and old Population I	Population II and old Population I	Mostly Population I
Percentage of observed galaxies	77%	*20%	3%

*This percentage does not include dwarf elliptical galaxies that are as yet too dim and distant to detect. Hence, the actual percentage of galaxies that are ellipticals is likely to be higher than shown here.

percluster contains dozens of individual clusters spread over a volume typically 150 million light-years across. **Figure 13-35** shows the distribution of superclusters in our vicinity of the universe. The Milky Way, along with the rest of the Local Group of galaxies, are members of the Laniakea Supercluster, identified in 2014. Laniakea is Hawaiian for ‘immeasurable heaven.’ The structure of this supercluster is still under investigation. Observations indicate that most superclusters are not gravitationally bound units—most clusters in each supercluster are drifting away from most of the other clusters in that same supercluster. Furthermore, the superclusters are all moving away from one another.

The distribution of superclusters appears to have the structure of a sponge How superclusters are distributed in space started to become clearer in the early 1980s, when astronomers began systematic surveys of the positions of observable objects in the nearby universe. Completed in 2005, the first Sloan Digital Sky Survey observed nearly 200 million galaxies, stars, and other celestial objects and measured the redshifts of more than a million of them. This survey, made in the northern hemisphere, covered about one-quarter of the sky and gives us a three-dimensional map of 100 times the volume of space as we had mapped before. Another set of observations, called the Two Degree Field (2dF) Survey, studied regions of both the northern and southern hemispheres.

These data reveal that most superclusters are located on the boundaries between enormous bubblelike *voids*



FIGURE 13-34 A Cluster of Galaxies This group of galaxies, called the Hercules cluster, is about 650 million light-years from Earth. Both elliptical and spiral galaxies within this cluster can be easily identified. (NOAO/AURA/NSF)

in which few galaxies are observed (Figure 13-35 and **Figure 13-36**). These voids are roughly spherical and typically measure between 100 million and 400 million light-years in diameter. They are not completely empty, however. Observations reveal hydrogen clouds in some of them, while others may be subdivided by strings of dim galaxies.

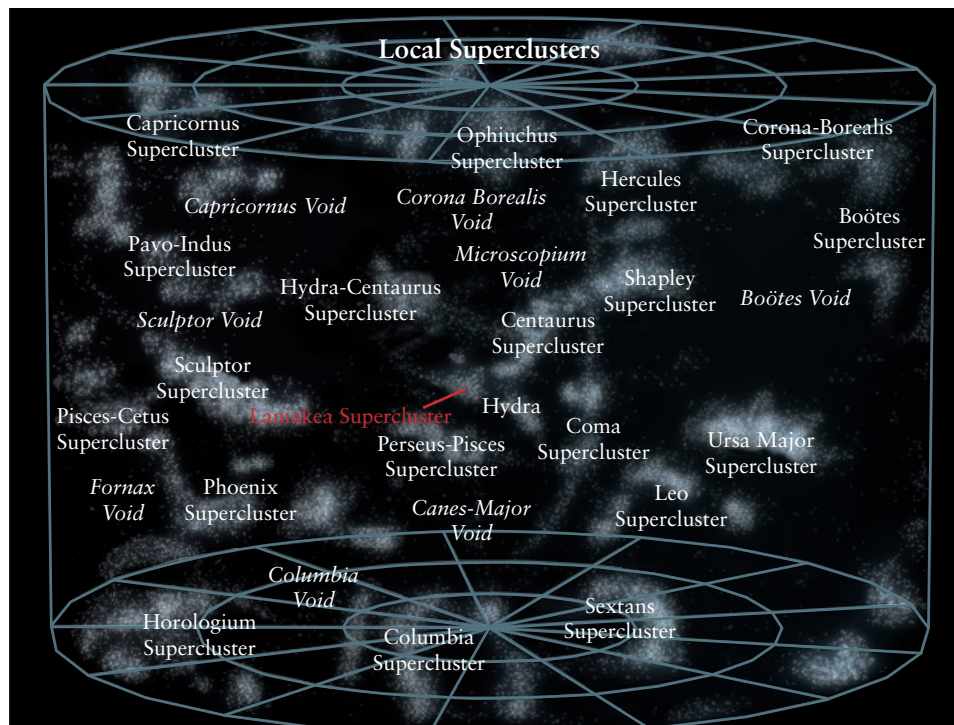


FIGURE 13-35 Superclusters in Our Neighborhood This diagram shows the distances and relative positions of superclusters within 950 million light-years of Earth. Note also the

labeling of some of the voids, which are large, relatively empty regions between superclusters. (Andrew Z. Colvin)

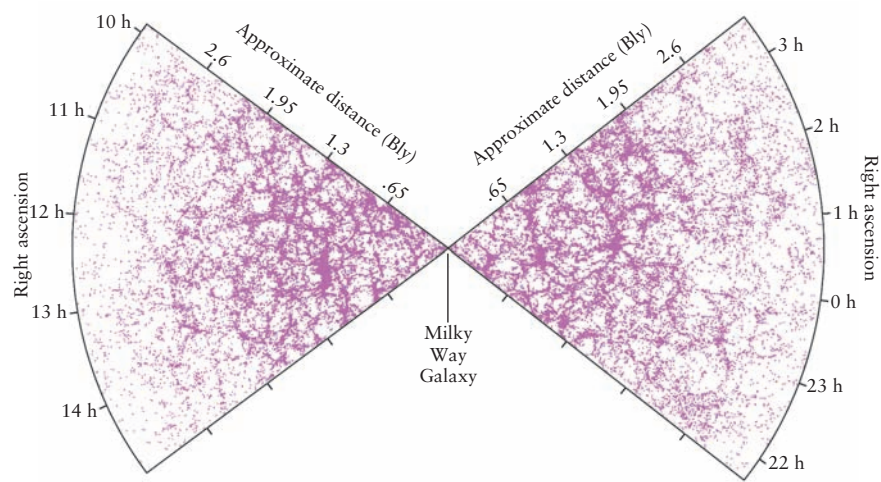


FIGURE 13-36 Structure in the Universe This map shows the distribution of 62,559 galaxies in two wedges extending in opposite directions from Earth out to distances of 3.25 Bly (billion light-years). For an explanation of right ascension (R.A.), see Section 1-3. Note the prominent voids surrounded by thin areas full of galaxies. (Courtesy of the 2dF Galaxy Redshift Survey Team)

The 2dF Galaxy Survey

Focus Question 13-11
Astronomers observe very hot (10^7 K) gas between clusters of galaxies. Why does this gas not appear in Figure 13-34?

The distribution of clusters of galaxies throughout the universe appears to have the structure of a sponge (Figure 13-37) or soap bubbles, with the galaxies located where the sponge or soapy material is, surrounded by voids. Astronomers believe that this spongy pattern contains important clues about conditions in the early universe that have not yet been fully fathomed.

Clusters of galaxies may appear densely or sparsely populated and regular or irregular in shape. A cluster of galaxies is said to be either a **poor cluster** or a **rich cluster**, depending on whether it contains less than or more than a thousand galaxies. For example, the Milky

Way Galaxy, the Andromeda Galaxy, and the Large and Small Magellanic Clouds belong to a poor cluster called the **Local Group**. The Local Group contains roughly 40 galaxies, over a third of which are dwarf ellipticals. Figure 13-38 shows a map covering most of the Local Group. The Virgo cluster is the nearest rich cluster, with more than 2000 galaxies.

Galaxies in the Local Group continue to be found. These discoveries happen primarily because astronomers are developing techniques to observe objects that lie in the same plane as, but beyond, the Milky Way. In addition to the 1994 discovery of the Sagittarius Dwarf Galaxy, the



FIGURE 13-37 Foamy Structure of the Universe A sponge structure recreates the distribution of bright clusters of galaxies throughout the universe. The empty spaces in the foam are analogous to the voids found throughout the universe. The spongy regions are analogous to the locations of most of the superclusters of galaxies. (© Image Source/SuperStock)

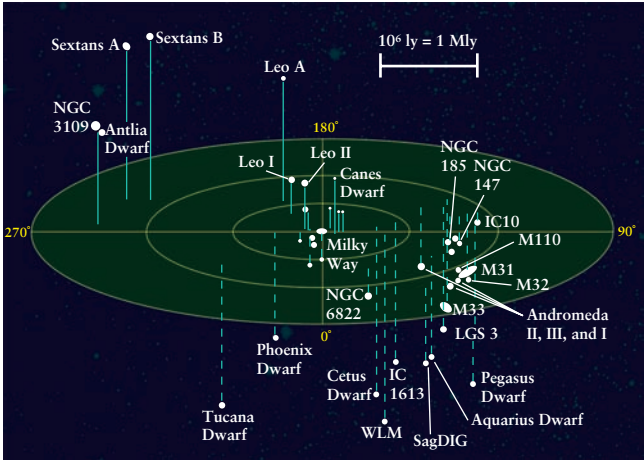


FIGURE 13-38 The Local Group Our Galaxy belongs to a poor, irregular cluster that consists of about 40 galaxies, called the Local Group. This map shows the distribution of about three-quarters of the galaxies. The Milky Way and Andromeda galaxies are the largest and most massive galaxies in the Local Group. Andromeda (M31) and the Milky Way are each surrounded by a dozen satellite galaxies. The recently discovered Canes Dwarf Galaxy is the Milky Way's nearest known neighbor.



R I V U X G

FIGURE 13-39 A Recently Discovered Member of the Local Group The galaxy Antlia was first detected in 1997. It lies about 3 million light-years away, outside the region depicted in Figure 13-38. This galaxy contains only about a million stars. (M. J. Irwin, Royal Greenwich Observatory, and A. B. Whiting and G. K. T. Hau, Institute of Astronomy, Cambridge University)

dwarf galaxy Antlia (Figure 13-39) was first observed in 1997 and the Canis Major Dwarf (see Figure 13-16) was found in 2003. The Canis Major Dwarf is only about 25,000 light-years from the Milky Way, which is closer to us than we are to the center of our Galaxy.

Astronomers also categorize clusters of galaxies according to their shapes. A **regular cluster** is distinctly spherical with a marked concentration of galaxies at its center. Numerous gravitational interactions over billions of years have spread the galaxies into this distribution. In contrast, galaxies in an **irregular cluster** are more randomly scattered about a sprawling region of the sky. The local group and the Virgo cluster are irregular.

Focus Question 13-12

What are two reasons that new galaxies are still being discovered in the Local Group?

The nearest example of a rich, regular cluster is the Coma cluster, located about 300 million light-years from us in the constellation of Coma Berenices (Figure 13-40). Despite its great distance, more than 1000 bright galaxies within it can be detected with our current telescopes.

Rich, regular clusters, like the Coma cluster, contain mostly elliptical and lenticular galaxies. Only 15% of the Coma cluster's presently known galaxies are spirals and irregulars. It is likely that the Coma cluster contains many thousands of as yet undetected dwarf ellipticals—maybe as many as 10,000 galaxies overall.

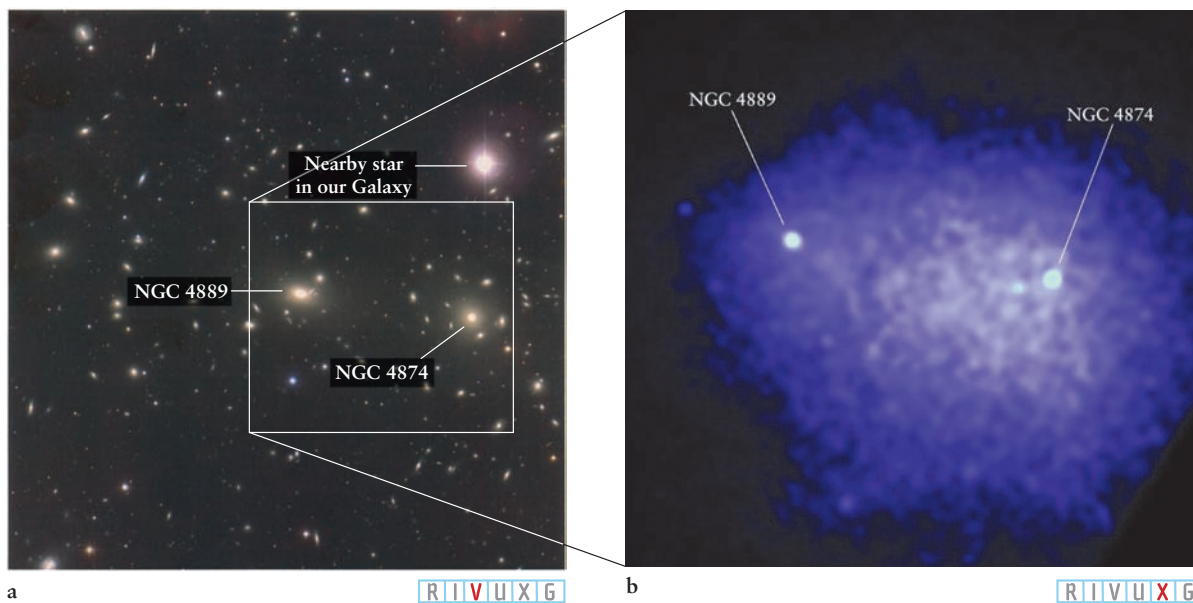


FIGURE 13-40 The Coma Cluster (a) This rich, regular cluster, containing thousands of galaxies, is about 300 million light-years from Earth. Only the inner region of the cluster is shown here. (b) Regular clusters are composed mostly of elliptical and lenticular galaxies and are sources of X-rays.

This Chandra image shows Coma's central region, which is 1.5 million light-years across. The gas cloud emitting most of these X-rays has a temperature of 100 million K. (a: Omar Lopez-Cruz & Ian Shelton/NOAO/AURA/NSF; b: NASA/CXC/SAO/A. Vikhlinin et al.)

Irregular clusters, such as the Virgo cluster and the Hercules cluster (see Figure 13-34), display a more even mixture of galaxy types. While two-thirds of the 200 brightest galaxies in the Hercules cluster are spirals, nearly a fifth are ellipticals, and the rest are irregular.

13-15 Galaxies in a cluster can collide and combine

As noted earlier, two or more galaxies in a cluster or two or more galaxies from neighboring clusters occasionally collide. Sometimes these galaxies pass through one another (Figure 13-41) and sometimes they merge. There is so much space between stars that the probability of two stars crashing into each other during galactic collisions is extremely small. On the other hand, the galaxies' huge clouds of interstellar gas and dust are so large that they do collide, slamming into each other and producing strong shock waves. The colliding interstellar clouds are stopped in their tracks.

A violent collision can strip both galaxies of their interstellar gas and dust. This happens because the stars in the two galaxies pass by each other and emerge as two separate galaxies, but the gas and dust from the two galaxies collide and stop each other. As a result, the gas and dust remain at the collision site as the stars depart. The energy associated with the collision heats this gas and dust to extremely high temperatures. This process is a major source of the hot **intergalactic gas** often observed in rich, regular clusters. The Chandra X-ray Observa-

tory has observed intergalactic gas with temperatures between 300,000 and 100 million K.

Another source of intergalactic gas is gas escaping from dwarf galaxies. This gas is ejected by supernovae of stars created in a burst of star formation when these galaxies formed. Whereas our Galaxy's gravitational force traps supernova remnants in the disk, this gas escapes into intergalactic space from the relatively weak gravitational attraction of dwarf galaxies.

Less violent collisions and near misses between galaxies allow the compressed interstellar gas and dust to cool sufficiently to have new stars form from it. These collisions can also stimulate galaxywide star formation, which may account for the **starburst galaxies** that blaze with the light of numerous newborn stars. These galaxies are characterized by bright centers surrounded by clouds of warm interstellar dust, indicating a recent, vigorous episode of star birth (Figure 13-42). Their warm dust is so abundant that starburst galaxies are among the most luminous objects in the universe at infrared wavelengths.

Observational evidence indicates that globular clusters are formed during collisions of interstellar gas and dust. While most globular clusters are over 10 billion years old, some galaxies also contain very young clusters and, like the Milky Way, have some very dense star-forming regions that are believed to be precursors of new globular clusters. Whether those in our Galaxy formed as a result of an early collision of the Milky Way with another galaxy has yet to be determined.

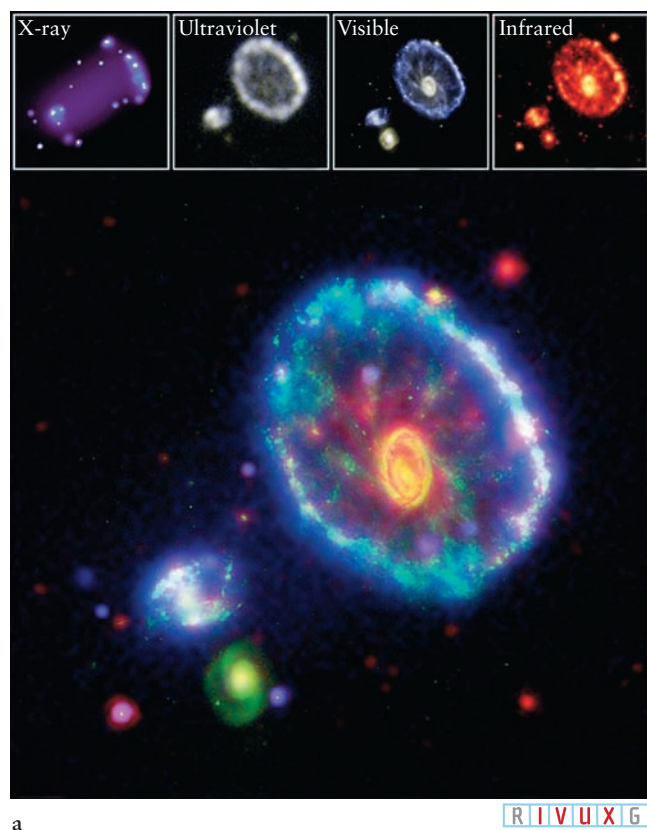


FIGURE 13-41 Galaxies with Rings (a) A composite image of the Cartwheel Galaxy. This ring-shaped assemblage 500 million light-years from Earth is likely the result of one galaxy, probably the blue-white one below it at the eight o'clock position, having passed through the middle of the larger one. Astronomers suspect that the passage created a circular density wave in the Cartwheel, analogous to the ripple created by a rock thrown in a pond. This wave stimulated a burst of star formation, creating many bright blue and white stars. Ultraviolet is in blue, visible light in green, infrared in red, and X-ray in violet. (b) Infrared image of the Andromeda Galaxy. The ring of hot dust indicates star formation, probably caused by the passage of another galaxy through Andromeda. The fact that the ring is disturbed suggests that yet another galaxy had a close interaction with Andromeda. (a: NASA/JPL-Caltech/P. Appleton [SSC/Caltech]; a1: NASA/CXC/A. Wolter & G. Trinchieri et al.; a2–a4: NASA/JPL-Caltech/P. Appleton [SSC/Caltech]; b: NASA/JPL-Caltech/K. Gordon [University of Arizona])

The Irr II starburst galaxy M82 is one member of a nearby cluster of about a dozen galaxies that include the beautiful spiral galaxy M81 and a fainter companion,

called NGC 3077 (Figure 13-43a). A radio survey of that region of the sky revealed enormous streams of hydrogen gas connecting these and other, smaller galaxies

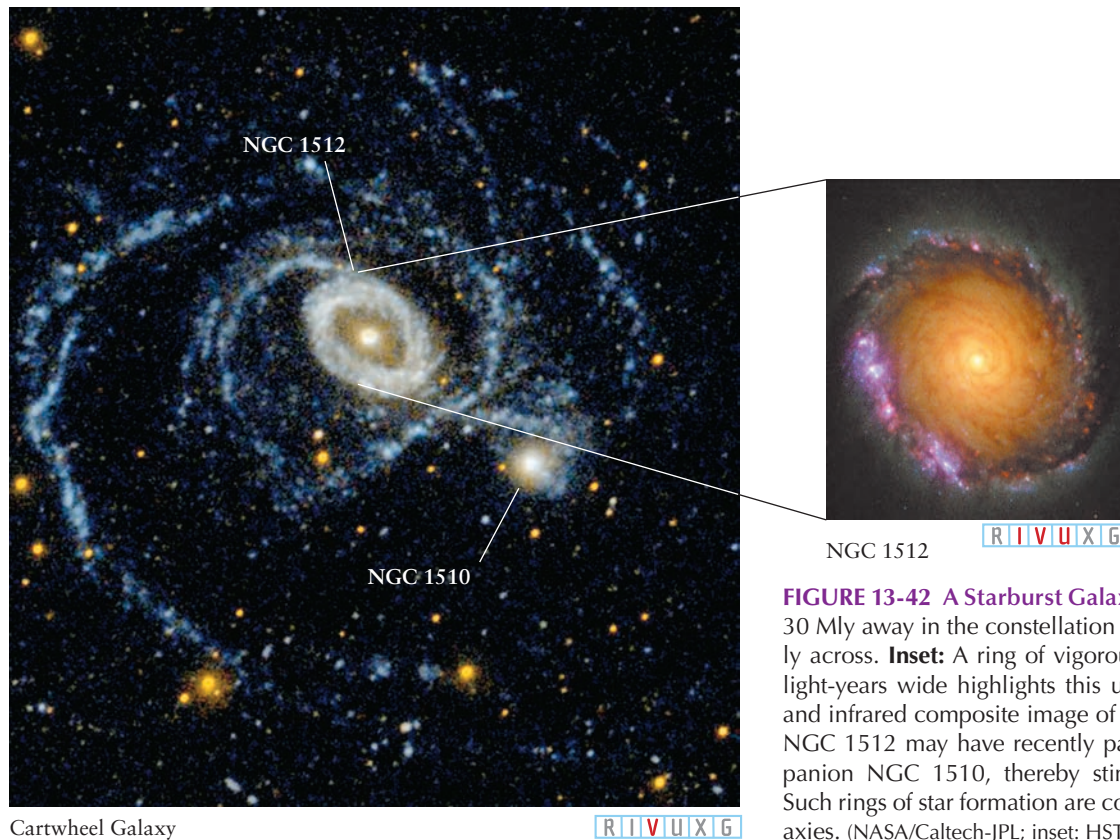


FIGURE 13-42 A Starburst Galaxy NGC 1512, located 30 Mly away in the constellation Horologium, is 70,000 ly across. **Inset:** A ring of vigorous star formation 2400 light-years wide highlights this ultraviolet, visible-light, and infrared composite image of the core of this galaxy. NGC 1512 may have recently passed close to its companion NGC 1510, thereby stimulating the starburst. Such rings of star formation are common in starburst galaxies. (NASA/Caltech-JPL; inset: HST/NASA/ESA)

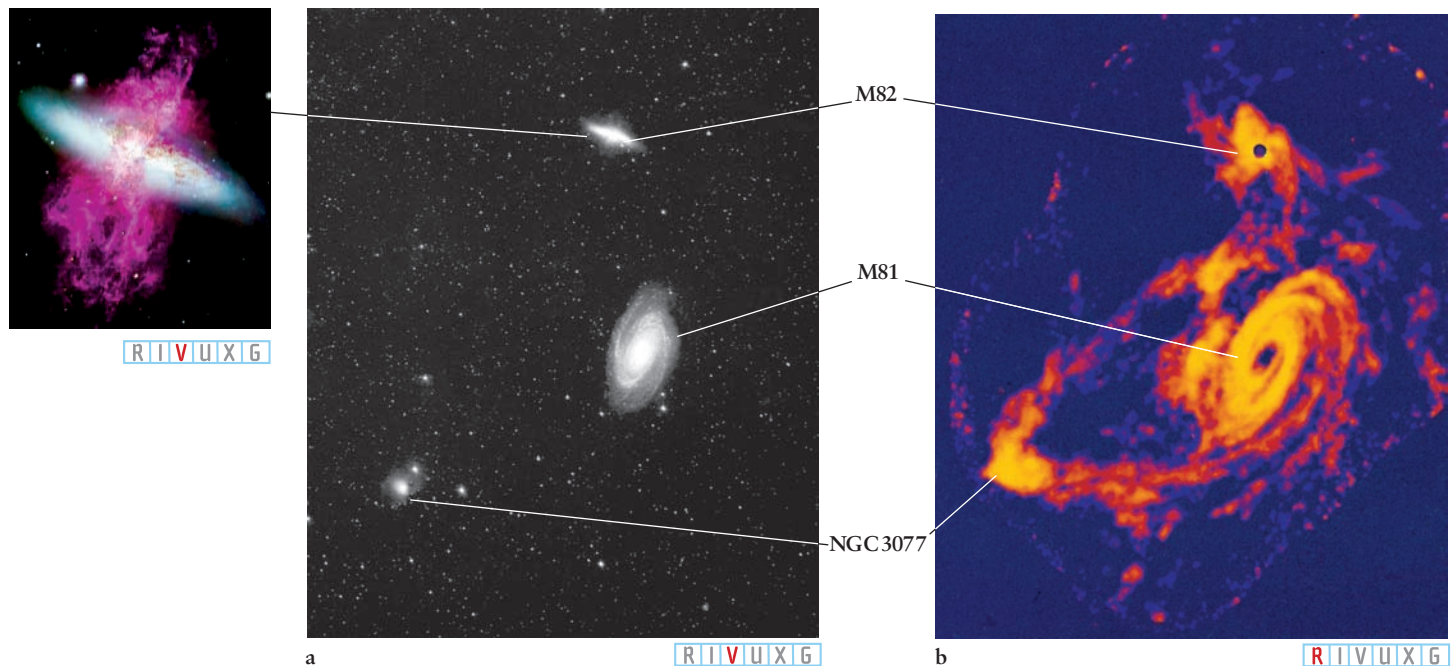


FIGURE 13-43 The M81 Group The Irr II starburst galaxy M82 is in a nearby cluster of about a dozen galaxies, including the spectacular spiral M81. Several of the galaxies in this cluster are connected by streamers of hydrogen gas. **(a)** The three brightest galaxies in the cluster at visual wavelengths. The inset on the left shows large volumes of hydrogen gas, in red, being ejected from M82.

(b) This radio image, created from data taken by the Very Large Array, shows the streamers of hydrogen gas that connect the bright galaxies and also several dim ones, seen as regions of bright orange here. (inset: Mark Westmoquette [University College London], Jay Gallagher [University of Wisconsin-Madison], Linda Smith [University College London], WIYN//NSF, NASA/ESA; a: Palomar Sky Survey; b: M. S. Yun, VLA, and Harvard)

in that cluster (Figure 13-43b). The loops and twists in these streamers suggest that the three galaxies have had several close encounters over the ages, most recently about 600 million years ago.

Gravitational interactions between colliding galaxies can also result in millions of stars being hurled out into intergalactic space along huge, arching streams, as seen in **Figure 13-44a**. In rich clusters, astronomers also observe near misses between galaxies (Figure 13-44b). If galaxies are surrounded by extended halos of dim stars, these near misses strip the galaxies of these outlying stars. In this way, a loosely dispersed sea of dim stars populates the space between galaxies in a cluster.

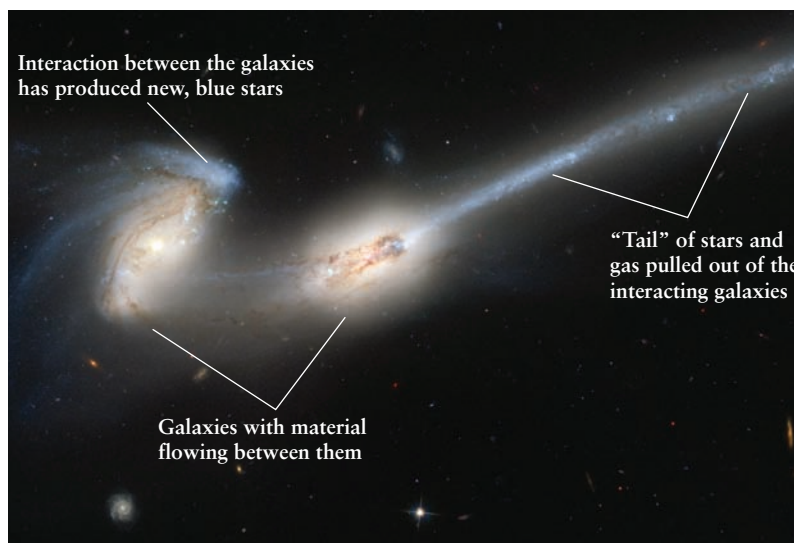
Similarly, after galaxies collide, some interior stars are flung far and wide, scattering material into intergalactic space. Such isolated stars have been observed by the Hubble Space Telescope. However, other stars slow down, often causing the remnants of the galaxies to merge. Several dramatic examples of **galactic mergers** have been discovered, such as the galaxy known as NGC 6240 (**Figure 13-45**). The supermassive black holes in the nuclei of the original two galaxies (see Section 12-15) have not yet merged (see Figure 13-45 inset) but are expected to do so within a few hundred million years. A pair of supermassive black holes was also observed in the nuclear region of the spiral galaxy NGC 3393, located some 160 Mly from Earth. They are also believed to be from the merger of two galaxies into one over a billion years ago.

As mentioned in Section 13-5 with regard to the Milky Way tearing apart the Canis Major Dwarf Galaxy, astronomers speak of **galactic cannibalism** when a large galaxy captures and “devours” a smaller one. Cannibalism differs from mergers in that the “dining” galaxy is significantly bigger than its “dinner,” whereas merging galaxies are about the same size.

Many astronomers suspect that giant ellipticals are the product of galactic mergers and cannibalism, including spiral–spiral collisions. Observations reveal that giant galaxies typically occupy the centers of rich clusters. Once formed, giant ellipticals can continue to grow because smaller galaxies are often located around them (see Figure 13-29). As they pass through the extended halo of a giant elliptical, these smaller galaxies slow down and are eventually cannibalized by the larger galaxy.

Figure 13-46, a computer simulation, shows a large, disk-shaped galaxy devour-

ing a small satellite galaxy. The large galaxy consists of 90% stars (blue) and 10% gas (white) by mass. It is surrounded by a halo of dark matter whose mass is about 3.3 times that of the disk. The satellite galaxy, which has a tenth of the mass of the large galaxy,



a NGC 4676



b IC 2163 (left) and NGC 2207

FIGURE 13-44 Interacting and Colliding Galaxies (a) Pairs of colliding galaxies often exhibit long “antennae” of stars ejected by the collision. This particular system is known as NGC 4676 or “the Mice” (because of its tails of stars and gas). It is 300 million light-years from Earth in the constellation Coma Berenices. The collision has stimulated a firestorm of new star formation, as can be seen in the bright blue regions. Mass can also be seen flowing between the two galaxies, which will eventually merge. (b) These two galaxies, NGC 2207 (right) and IC 2163, are orbiting and tidally distorting each other. Their most recent close encounter occurred 40 million years ago when the two were perpendicular to each other and about 1 galactic diameter apart. Computer simulations indicate that they should eventually coalesce. (a: NASA, H. Ford [JHU], G. Illingworth UCSC/Lick, M. Clampin [STScI], G. Hartig [STScI]/ACS Science Team/ESA; b: NASA/ESA and the Hubble Heritage Team [STScI])

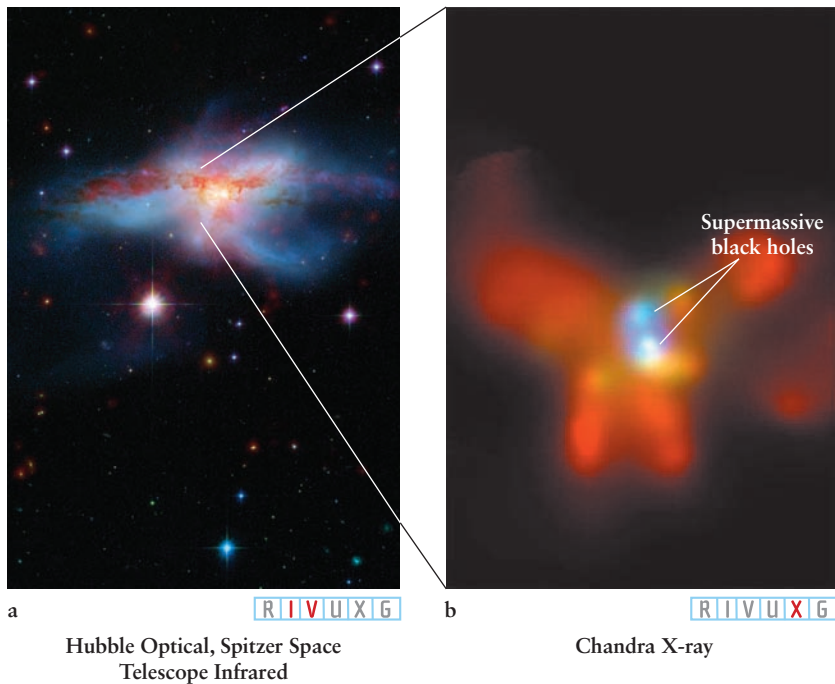


FIGURE 13-45 Merging Galaxies This object, NGC 6240, in the constellation Ophiuchus, is the result of two spiral galaxies in the process of merging. The widespread blue-green area reveals that the collision between the two galaxies has triggered an immense burst of star formation. **Inset:** The Chandra X-ray Observatory shows that at the heart of this system are two supermassive black holes, one from each of the original galaxies. Within a few hundred million years, these black holes are expected to merge into a single more massive black hole. (a: NASA/JPL-Caltech/STScI ESA/S. Bush et al. [Harvard-Smithsonian CfA]; b: S. Komossa & G. Hasinger [MPE] et al., CXC, NASA)

contains only stars (yellow). Initially, the satellite is in circular orbit about the large galaxy. Note that spiral arms appear in the large galaxy as the collision proceeds. Two billion years elapse as the satellite spirals in toward the core of the large galaxy. Although much material is stripped from the satellite, most of its stars plunge into the central region of the large galaxy.

13-16 Dark matter helps hold together individual galaxies and clusters of galaxies

Quite a bit of evidence has accumulated to support the idea of extended dark matter halos surrounding all galaxies. For example, many spiral

Focus Question 13-13

After the Milky Way collides with the Andromeda Galaxy, in what type of galaxy will our descendants likely live?

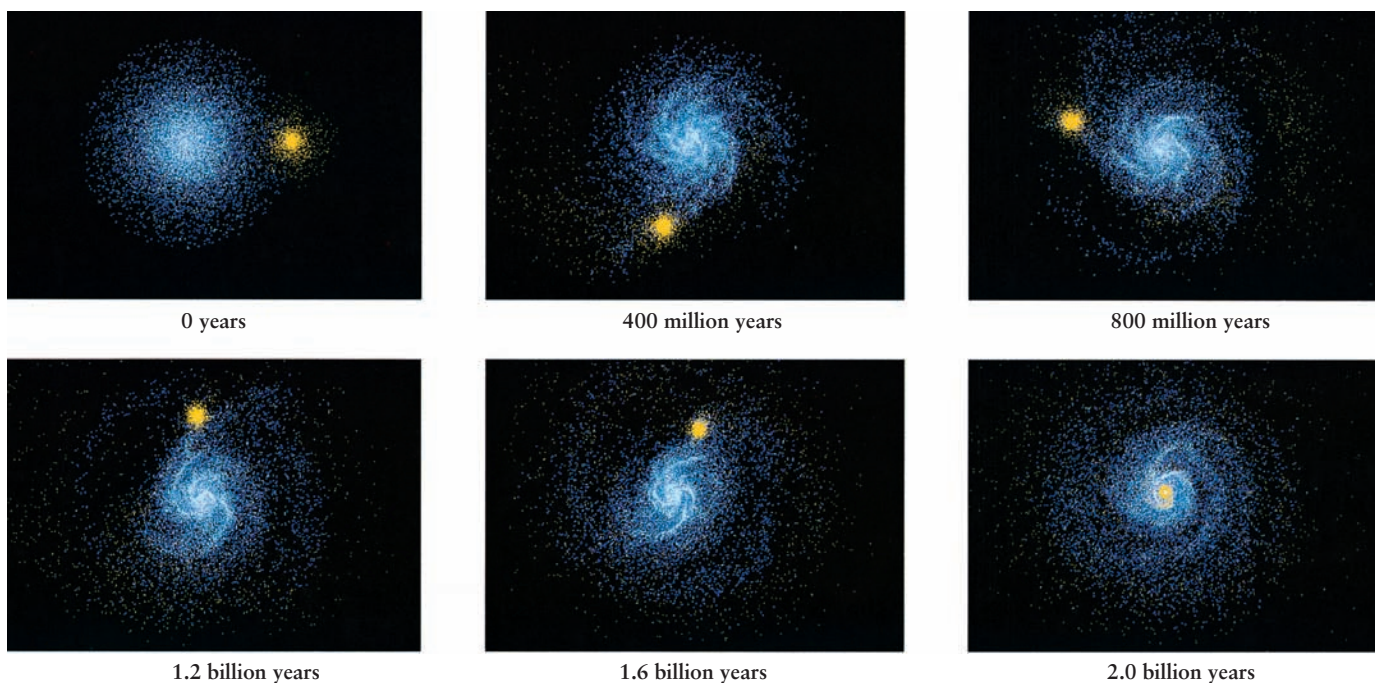


FIGURE 13-46 Simulated Galactic Cannibalism This computer simulation shows a small galaxy (yellow stars) being devoured by a larger, disk-shaped one (blue stars, white gas). Note how

spiral arms are generated in the disk galaxy by its interaction with the satellite galaxy. (Lars Hernquist, Institute for Advanced Study with simulations performed at the Pittsburgh Supercomputing Center)

galaxies have rotation curves similar to those of our Milky Way (recall Figure 13-18). These rotation curves (Figure 13-47) remain remarkably flat out to surprisingly great distances from the galaxies' centers. In all cases, the orbital speed is fairly constant. According to Kepler's third law, we should see a decline in orbital speed toward the outer portions of a galaxy. In most cases, we do not, which implies that we still have not detected the true edge of these and many similar galaxies. As with the Milky Way Galaxy, astronomers conclude that a considerable amount of dark matter, introduced in Section 13-7, must extend well beyond the visible portions of most galaxies. While the fraction of a galaxy's mass in dark matter is typically about 80%, some galaxies have much more of it. In 2010, a nearby dwarf galaxy, Segue 1, was observed to have a few thousand stars, but they are orbiting so rapidly that they must be under the influence of 500 times as much mass in dark matter!

As with the stars orbiting in a galaxy, there must be sufficient matter to provide the gravitational force necessary to bind galaxies together in clusters. However, *no cluster of galaxies contains enough visible matter to stay bound together*. Most of the matter binding clusters of galaxies together is dark matter. This must be true because, otherwise, the galaxies would have long ago wandered apart in random directions, and the clusters would no longer exist.

Analyses of all the visible mass demonstrate that the total mass needed to bind together a typical rich cluster is about 10 times greater than the mass of matter we can observe. Furthermore, observations of the distribution of galaxies in clusters have determined that the dark matter usually has the same distribution as the galaxies

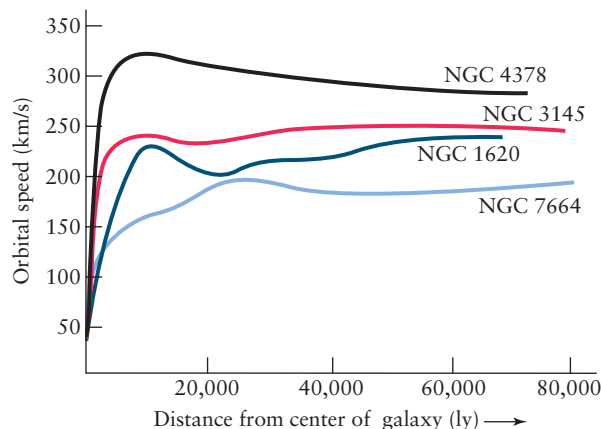


FIGURE 13-47 The Rotation Curves of Four Spiral Galaxies This graph shows how the orbital speed of material in the disks of four spiral galaxies varies with the distance from the center of each galaxy. If most of each galaxy's mass were concentrated near its center, these curves would fall off at large distances. But these and many other galaxies have flat rotation curves that do not fall off. This indicates the presence of extended halos of dark matter. See Figure 13-18 to compare these to the Milky Way's rotation curve.

in the cluster and that each galaxy corresponds to a concentration of dark matter. We will see some exceptions to this latter point later in this chapter.

Astronomers using X-ray telescopes have discovered that some of the dark matter is actually very hot intergalactic gas. Satellite observations have revealed X-rays pouring from the space between galaxies in rich clusters that does not emit discernible visible light. This X-ray radiation is emitted by substantial amounts of hot intergalactic gas, as noted in the previous section. Figure 13-40b shows this hot gas in the center of the Coma cluster. Observations by the Chandra X-ray Observatory indicate that this intergalactic gas contains twice as much mass as exists in all of that cluster's galaxies. Because its physical nature has now been identified, this intergalactic matter is no longer considered to be "dark."

Nature assists astronomers with visualizing dark matter by the way its mass gravitationally warps space, as described by Einstein's theory of general relativity (see Section 12-11). We saw in Section 13-7 how the same effect occurs on small scales here in the Milky Way. The local distortion is called *microlensing*, and when dark matter and the visible matter in entire galaxies or clusters of galaxies focus light from more distant galaxies, the effect is called **gravitational lensing** (Figure 13-48a). Figures 13-48b and 13-48c show the gravitationally lensed images of distant galaxies caused by dark matter and closer galaxies.

In 2000, astronomers observed an intergalactic cloud of gas the size of a galaxy. Observations revealed that for that gas to remain bound together, it must contain about 100 billion solar masses of matter. Because no galaxy exists in that space, astronomers conclude that the gas is held together by that much dark matter.

Sometimes visible matter and dark matter separate, shown in Figure 13-48d. This is presumably the result of two clusters of galaxies colliding (Figure 13-48e). The details of the mechanism by which this separation occurs are still under investigation. In 2007, astronomers observed distant galaxies undergoing gravitational lensing without the presence of an intervening visible galaxy. They deduced that the effect was caused by just the presence of dark matter. By observing such gravitational lensing, astronomers have observed a filament of dark matter stretching between two galaxy clusters, Abell 222 and Abell 223 (Figure 13-48f). This filament spans 59 Mly.

The amount of gravitational lensing by dark matter alone or by dark matter and a galaxy or by a cluster of galaxies provides astronomers with more than just enlarged, distorted images of distant objects. It tells us how much mass is contained in the dark matter and galaxy or cluster that does the lensing. The greater the mass doing the lensing, the greater the effect on the distant object. The visible matter in the galaxies accounts for only about 10% to 20% of the gravitational lensing that occurs. Thus, dark matter accounts for 80% to 90% of the mass in the universe.

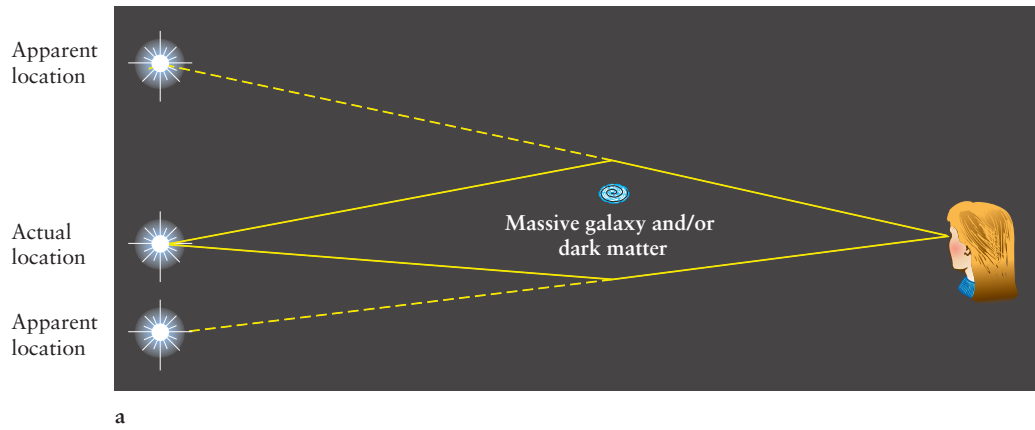
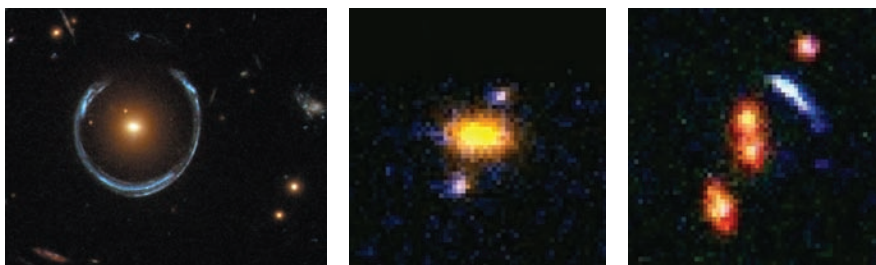


FIGURE 13-48 Gravitational Lensing of Extremely Distant Galaxies

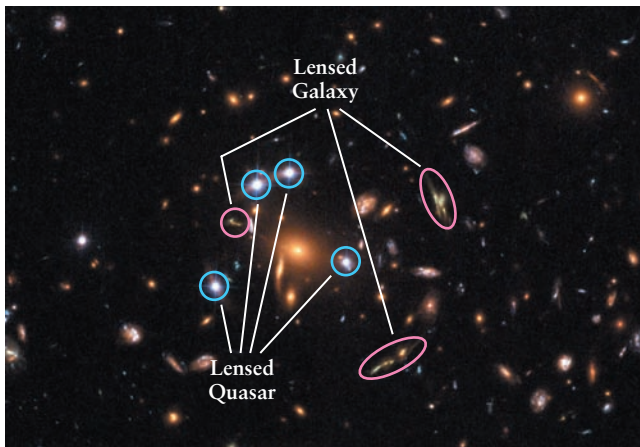
(a) Schematic of how a gravitational lens works. Light from the distant object changes direction due to the gravitational attraction of the intervening galaxy and underlying dark matter. The more distant galaxy appears in different places than it actually is. **(b)** Three examples of gravitational lensing: (1) The blue ring is a galaxy that has been lensed by the redder elliptical galaxy; (2) a pair of bluish images of the same object lensed symmetrically by the brighter, redder galaxy between them; and (3) the lensed object appears as a blue arc under the gravitational influence of the group of four galaxies.

(c) This image shows the gravitational lensing of a distant galaxy and a distant quasar by a cluster of galaxies between them and us. **(d)** This cluster contains over a thousand galaxies. The red shows the cluster's hot gas, while the blue is its dark matter, determined by gravitational lensing. **(e)** A mechanism proposed to explain the separation of gas and dark matter in clusters like (d). **(f)** Gravitational lensing by this filament of dark matter between two clusters of galaxies reveals its presence. (b1: ESA/Hubble and NASA; b2 and b3: NASA, ESA, A. Bloton [Harvard-Smithsonian CfA]; c: NASA, ESA, K. Sharon [Tel Aviv University], and E. Ofek [Caltech]; d: X-ray: NASA/CXC/CfA/M. Markevitch et al.; visible: NASA/STScI, Magellan/U. Arizona/D. Clew et al.; lensing map: NASA/STScI; e: NASA/CXC/M. Weiss; f: Jörg Dietrich, University of Michigan/University Observatory Munich)



b

R I V U X G



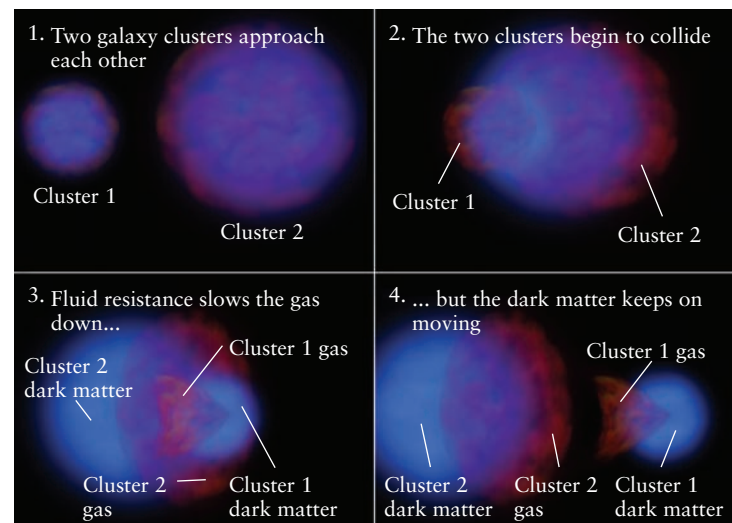
c



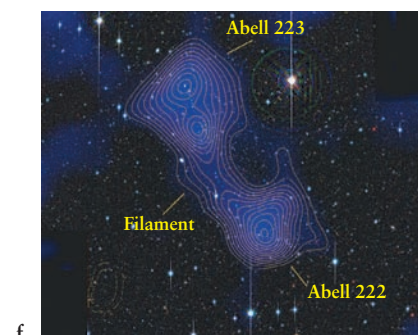
d

Composite image of galaxy cluster 1E0657-56 showing visible galaxies, X-ray emitting gas (red), and dark matter (blue).

R I V U X G



e A model of how the gas and dark matter in 1E0657-56 could have become separated



f

SUPERCLUSTERS IN MOTION

Whenever an astronomer finds an object in the sky, one of the first tasks is to determine its composition. As we saw in Chapters 3 and 4, that means attaching a spectrograph to a telescope and recording the object's spectrum. As long ago as 1914, the American astronomer Vesto Slipher, working at the Lowell Observatory in Arizona, took spectra of “spiral nebulae,” now known to be spiral galaxies. He was surprised to discover that the spectral lines of 11 of the 15 spiral nebulae that he studied were substantially redshifted. These redshifts indicate that they are moving away from the Milky Way at significant speeds. This result left scientists at that time with a major question: Is the universe expanding?

Focus Question 13-14

If the dark matter in Figure 13-48d suddenly vanished, what would we see of the clusters of galaxies behind it?

13-17 The redshifts of superclusters indicate that the universe is indeed expanding

During the 1920s, Edwin Hubble and Milton Humason recorded the spectra of many galaxies with the 100-in. telescope on Mount Wilson, confirming that most galaxies are rapidly receding from the Milky Way. This recession implies that the universe *is* expanding, although, as we will see in a moment, it does not imply that the Milky Way is at the center of it!

7 Using the Doppler effect (recall Section 3-2), Hubble calculated the speed at which the galaxies are moving away from us. Using techniques such as determining the brightness of Cepheid variables (see Section 13-1), he also estimated the distances to several of these galaxies. He found a direct correlation between the distance to a galaxy and the size of its redshift: *Galaxies in distant superclusters are moving away from us more rapidly than galaxies in nearby superclusters.* Figure 13-49 shows this correlation for five elliptical galaxies. This recessional motion pervades the universe and is now called the **Hubble flow**.

The separation of clusters in the same supercluster from each other is slower than the general Hubble flow of the universe because these clusters are relatively close to one another, so their gravitational attractions slow each other down significantly. Keep in mind that the Hubble flow does not occur for galaxies in any given cluster, because all of those galaxies are gravitationally bound to each other. To recap, the Hubble flow applies to the separation of superclusters from one another.

When Hubble plotted the redshift data on a graph of distance versus velocity, he found that the points lie roughly along a straight line. The galaxies he observed were relatively close to the Milky Way (representing a small fraction of the volume of the universe). To get the big picture of how the universe is expanding, astro-

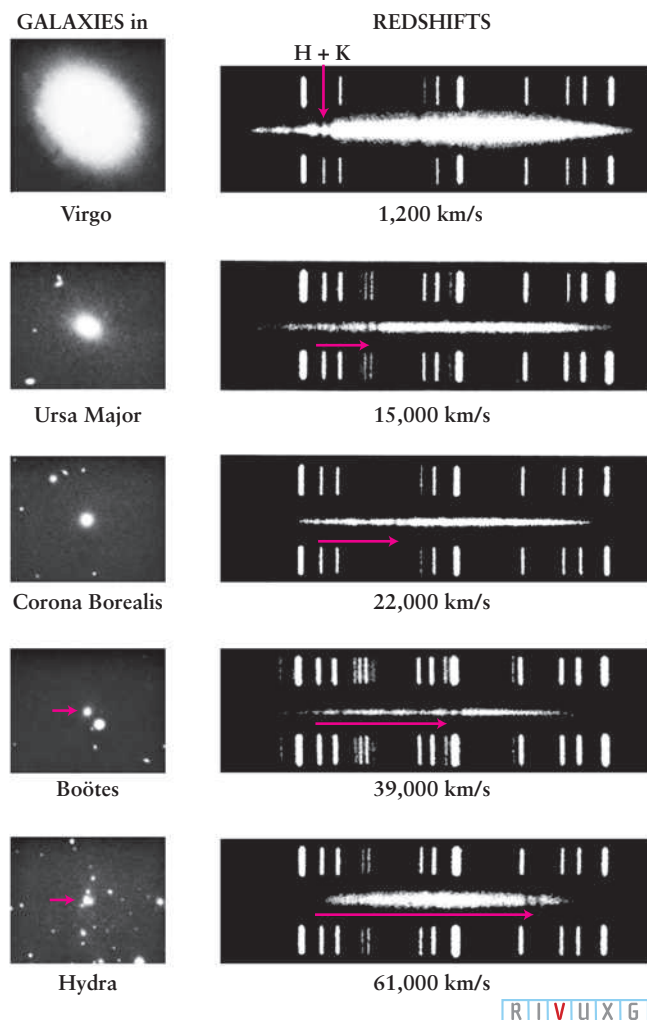


FIGURE 13-49 Five Galaxies and Their Spectra The photographs of these five elliptical galaxies were all taken at the same magnification. They are labeled according to the constellation in which each galaxy is located. The spectrum of each galaxy is the hazy band between the comparison spectra at the top and bottom of each plate. In all five cases, the so-called H and K lines of calcium are seen. The recessional velocity (calculated from the Doppler shifts of the H and K lines) appears below each spectrum. Note that the fainter—and thus more distant—a galaxy is, the greater is its redshift. (Carnegie Observatories)

ners must measure the redshifts and distances to galaxies much further away. Although redshift measurements from spectra can be quite precise, it is very difficult to accurately measure the distances to remote galaxies. We cannot use the method of stellar parallax at the distances of galaxies. Recall from Section 10-1 that only the distances to the stars in our Galaxy within about 490 pc (1600 ly) can be determined very precisely this way.

Distances to other galaxies are determined from observations of the apparent magnitudes of objects in them or apparent magnitudes of entire galaxies.

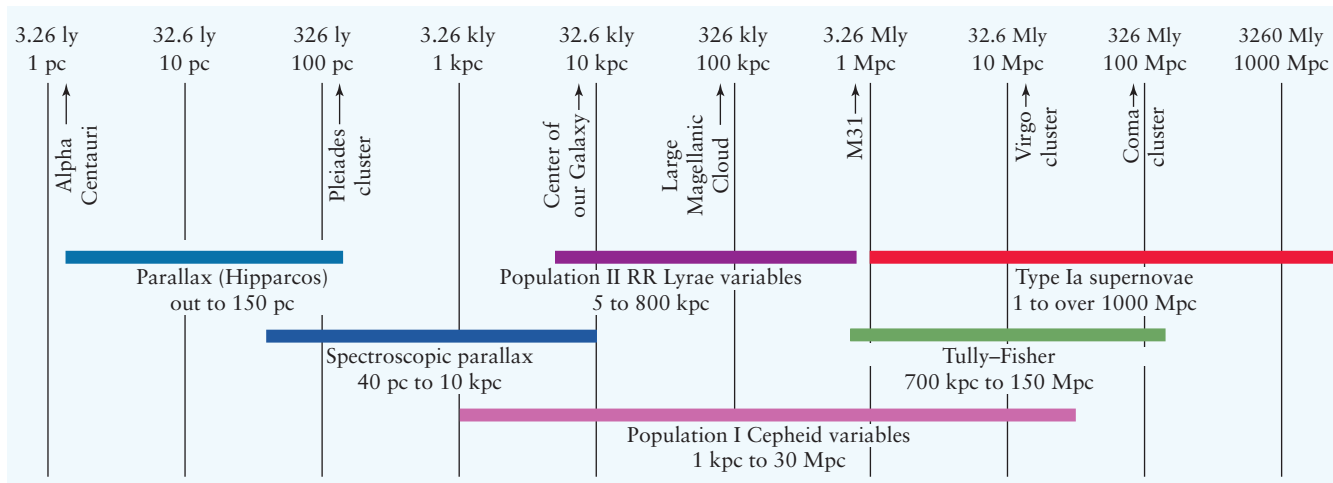


FIGURE 13-50 Techniques for Measuring Cosmological Distances Astronomers use different methods to determine

different distances in the universe. All of the methods shown here are discussed in the text.

These observations are compared to known absolute magnitudes for similar objects. Having the apparent magnitude and the absolute magnitude enables astronomers to calculate distances, as discussed in Discovery 10-2: The Distance–Magnitude Relationship. Astronomers use the term **standard candle** to denote any object whose absolute magnitude is known. The various techniques used to determine large-scale distances are presented in **Discovery 13-1: The Tully–Fisher Relation and Other Distance-Measuring Techniques**. **Figure 13-50** summarizes the distances for which all of the measuring techniques described here are valid.

Figure 13-51 shows the relationship between the distances to galaxies and their recessional motion, including many whose extreme distances from us were measured using the apparent magnitude of Type Ia supernovae in them. You can see that all the distance–redshift data represent a straight line. The information in this figure can therefore be stated as a simple formula, called the **Hubble law**:

$$\text{Recessional velocity} = H_0 \times \text{distance}$$

The straight line in **Figure 13-51** has a slope denoted by the constant H_0 (pronounced “H naught”), commonly called the **Hubble constant**. Using this diagram, astronomers have calculated that the Hubble constant is

$$H_0 = 69 \pm 2 \text{ km/s/Mpc}$$

This equation and its constant, H_0 , tell us that for every increase in distance of 1 Mpc, the galaxies are moving away from us 69 km/s faster than are those galaxies 1 Mpc closer to us. The Mpc is the standard unit of large distances used in this equation. The plus or minus 2 km/s/Mpc indicates the possible range for H_0 ,

when taking into account all of the errors that may still remain in the observations and calculations. A significant source of error is the effect of intergalactic gas that causes the standard candles to appear dimmer than they would if the gas were not present. Astronomers must determine the amounts of both intergalactic and interstellar gas and take their effects into account, which is often hard to do. It is worth noting that a variety of data sets used to determine the Hubble constant recently have given results ranging from $68 < H_0 < 71$, so clearly finding H_0 is still a work in progress.

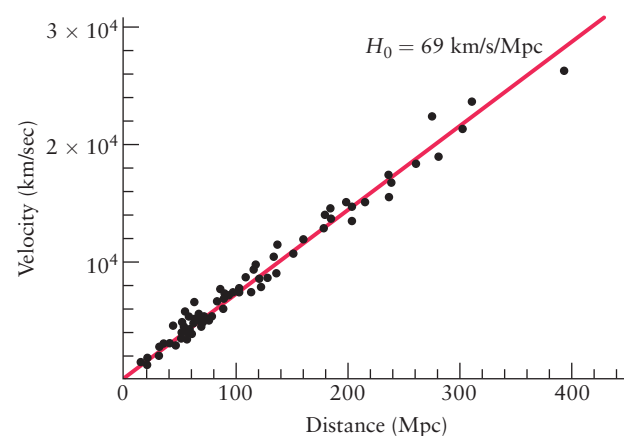


FIGURE 13-51 The Hubble Law The distances and recessional velocities of distant galaxies are plotted on this graph. The straight line is the “best fit” for the data. This linear relationship between distance and speed is called the Hubble law. For historical reasons, distances between galaxies, clusters of galaxies, and superclusters of galaxies are usually given in megaparsecs (Mpc) rather than millions of light-years.

DISCOVERY 13-1

The Tully–Fisher Relation and Other Distance-Measuring Techniques

Spectroscopic parallax (see Section 10-9) provides distances for stars up to 10 kpc (33 kly, where kly denotes 1000 light-years). This distance is still within the Milky Way. To determine distances to other galaxies, we need sources in them that become very bright compared to the luminosity of normal stars and whose absolute magnitudes are well known. Using the brightnesses of RR Lyrae variable stars (see Section 11-11), astronomers can measure distances to the Large and Small Magellanic Clouds, about 100 kpc (330 kly). To measure the Hubble flow, we must determine distances to much more distant galaxies.

Cepheid variable stars (see Section 13-2 and Figure 13-3) can now be observed out to 30 Mpc (100 Mly, where Mly denotes a million light-years) from Earth. The distances to galaxies in this nearby volume of space can thus be determined from the Cepheid period–luminosity law. Plotting the distances to galaxies versus their recessional velocity for the small volume of space extending out 30 Mpc from Earth (shown on the left side of the graph in Figure 13-51) enabled astronomers to get a first estimate for H_0 from the slope of the line through these data. This result gives $H_0 \approx 75$ km/s/Mpc. Errors in both distance and recessional velocities led to inaccuracies in the slope of this curve.

Beyond 30 Mpc, even the brightest Cepheid variables, which have absolute magnitudes of about -6 , are not visible with current technology. In the 1970s, the astronomers Brent Tully and Richard Fisher developed another method for determining distances. They discovered that the width of the hydrogen 21-cm emission line of a spiral galaxy (discussed in Section 13-3) is related to the galaxy’s absolute magnitude: *The broader the line is, the brighter the galaxy.* The connection goes like this: Interstellar gas clouds in each galaxy orbit with speeds depending on their locations in the galaxy, as we have seen. Therefore, clouds in different places in a galaxy have different speeds toward or away from us, meaning that they have different Doppler shifts as seen from Earth. Consequently, we observe a variety of slightly different wavelengths from these clouds, centered around

21 cm: This combination of emissions makes the 21-cm emission line appear “broad.”

The **Tully–Fisher relation** says that the greater the galaxy’s mass is, the greater is its luminosity and the more rapidly the stars and gas orbit in it. The faster it rotates, the greater the range of speeds and, hence, the greater the range of Doppler shifts we see from it. Measuring the width of the 21-cm line, therefore, tells us the galaxy’s mass. Assuming that stars with different masses and, hence, luminosities occur with the same frequencies in each galaxy, the galaxy’s mass tells us its *absolute* magnitude. Combining this with observations of the galaxy’s *apparent* magnitude allows us to calculate its distance from us using the distance–magnitude relationship (see Discovery 10-2: The Distance–Magnitude Relationship).

Because line widths can be measured quite accurately, astronomers can use the Tully–Fisher relation to determine the luminosities (absolute magnitudes) of many spiral galaxies and thus their distances from Earth.

The key method for determining the distances to the most remote galaxies is to measure the brightness of Type Ia supernova explosions within them. These supernovae all reach an absolute magnitude of -19 at the peak of their outbursts (see the figure below). Using their observed apparent magnitudes, their distances can be calculated. With this method, astronomers in 2001 measured a supernova nearly 3 billion parsecs (10 billion light-years) away, and more distant Type Ia supernovae have been measured since then. Measurements of distances to galaxies using the Tully–Fisher relation and Type Ia supernovae added data to Figure 13-51. Several space telescopes, most notably the Wilkinson Microwave Anisotropy Probe (WMAP) and the Planck Space Telescope, yield more data. Combining the data from the different methods yields a Hubble constant of 69 km/s/Mpc.

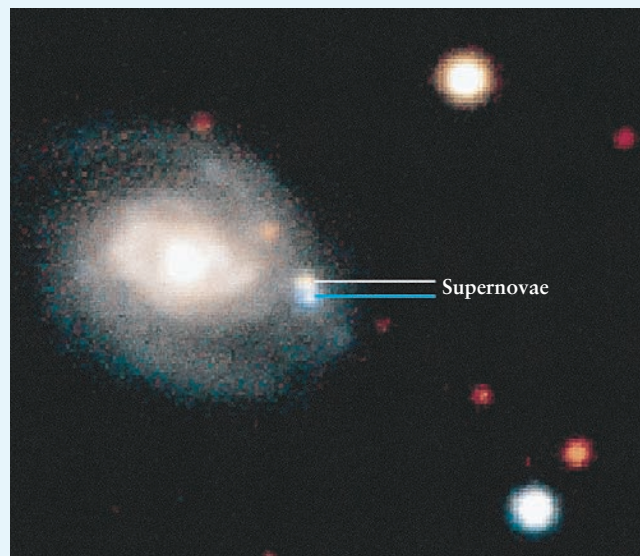


FIGURE D13-1 Two Supernovae in NGC 664 In 1997, the rare occurrence of two supernovae in the same galaxy at the same time was observed in the spiral galaxy NGC 664, located about 300 Mly (90 Mpc) from Earth. Supernovae observed in remote galaxies are important standard candles used by astronomers to determine the distances to these faraway objects. The two supernovae overlap each other, as shown. The upper, yellow-orange supernova was observed to occur 2 months before the hotter, blue one, which was observed to occur less than 2 weeks before this image was made and had not yet achieved maximum brightness. (Perry Berlind and Peter Garnavich, Harvard Smithsonian Center for Astrophysics)

DISCOVERY 13-2

The Expanding Universe

To understand the universe, scientists build models—mathematical representations of the world around us. In developing these pictures, it often helps to begin with a simple analogy. Because the expansion of the universe is hard to visualize, imagine for a moment that batter for a chocolate-chip cake is floating in an oven in the International Space Station. As the cake bakes, it expands in all directions and the chocolate chips move farther apart. The chips do not get larger and do not move through the batter. Each chocolate chip remains at rest in its own little bit of cake and is carried along as the cake spreads out.

Now, think of each chocolate chip as a supercluster of galaxies and the batter between the chocolate chips as the rest of spacetime. As the universe expands, the distance between widely separated superclusters of galaxies grows larger and larger. The expansion of the universe is the expansion of spacetime.

Suppose you were on one particular chocolate chip, say, the chocolate chip labeled 3 in the oven at the left in the figure. You would see chocolate chips 2 and 4 moving away from you with equal velocities. Every 10 minutes,

chocolate chips 2 and 4 get 1 cm farther away as the cake expands, as shown in the oven on the right. In that same time interval, chocolate chips 1 and 5 move twice as far away from you as chocolate chips 2 and 4. To go twice as far away in the same time, chocolate chips 1 and 5 must be moving twice as fast as the closer chocolate chips 2 and 4. This is exactly what Hubble's law of cosmological expansion (recessional velocity = $H_0 \times \text{distance}$) says: Double the distance and you double the velocity.

To see why Earth need not occupy the center of the universe, put yourself on chocolate chip 4. You see chocolate chips 3 and 5 moving away at the same rate—in fact, they recede at the same rate as you saw chocolate chips 2 and 4 move away when you were on chocolate chip 3. From chocolate chip 4, chocolate chips 2 and 6 are moving away twice as fast as chocolate chips 3 and 5. From any chocolate chip, you will see all of the other chocolate chips moving away. Similarly, no matter where we are in the universe, we would see all of the other superclusters receding from us in the same manner.

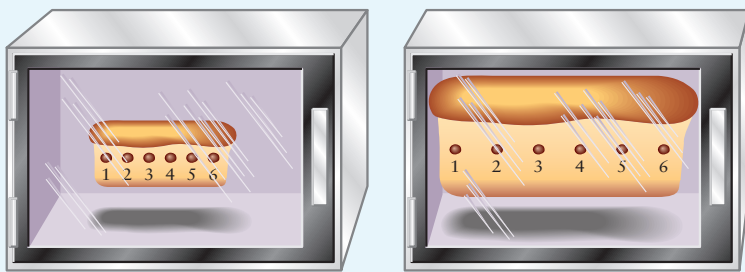


FIGURE D13-2 The Expanding Chocolate-Chip Cake Analogy The expanding universe can be compared to a chocolate-chip cake baking and expanding in the International Space Station. Just as all the chocolate chips move apart as the cake rises, the superclusters of galaxies recede from each other as the universe expands.

The relationship between the distances to galaxies and their redshifts was one of the most important astronomical discoveries of the twentieth century. It tells us that we are living in an expanding universe, and the Hubble law reveals the speed of that expansion. It also shows why we are not in the center of the universe, even though the other superclusters are all receding from us, as discussed in **Discovery 13-2: The Expanding Universe**.

13-18 Astronomers are looking back to a time when galaxies were first forming

The Hubble Space Telescope has been looking deeper and deeper into the universe. In 1998, working in concert with the Keck I telescope, Hubble observed galaxies 8 Bly (where Bly denotes a billion light-years) away, many of them in the process of merging (**Figure 13-52**). The same year it saw galaxies nearly 12 Bly away (**Figure 13-52b**). In 2012, galaxies up to 13.5 Bly away were observed. We are therefore seeing these latter objects as they were less than a billion years after the universe came into existence.

Although many of the early galaxies (meaning the ones farthest away from us) are not as well formed as nearby galaxies, the existence of galaxy-sized collections of stars so far back in time is an important clue to the processes of star and galaxy

Focus Question 13-15

If the Hubble constant were 3 times its present value, how much slower or faster would superclusters be moving apart?

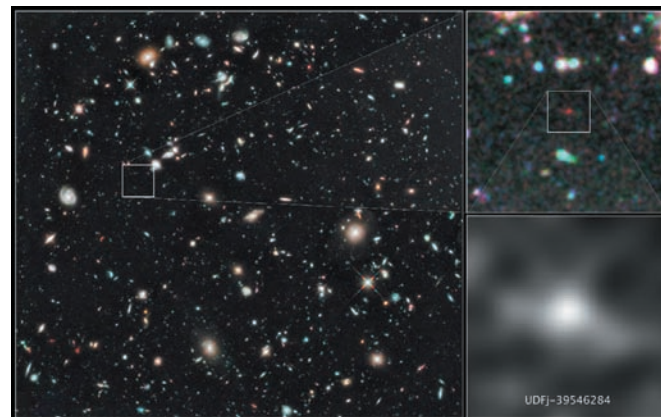
Insight Into Science

Power in Numbers Much of the power of science comes from expressing systematic data in mathematical form. By themselves, the spectra of galaxies give a qualitative picture of galactic motion, and the pattern of their recessional velocities suggests that a large-scale property of the universe is in effect here. When the data are converted into the equation, called the *Hubble law*, it becomes clear that the universe is expanding, and we can begin to look for the reason behind this expansion.



a

FIGURE 13-52 Distant Galaxies (a) The young cluster of galaxies MS1054-03, shown on the left, contains many orbiting pairs of galaxies, as well as remnants of recent galaxy collisions. Several of these systems are shown at the right. This cluster is located 8 billion light-years away from Earth. (b) This image of more than 300 spiral, elliptical, and irregular galaxies contains a galaxy that is an estimated 13.2



b

billion light-years from Earth. A compact galaxy of blue stars that existed 480 million years after the Big Bang is shown in the images on the right. (a: Pieter van Dokkum [University of Groningen], ESA, and NASA; b: NASA, ESA, G. Illingworth, and R. Bouwens [University of California, Santa Cruz], and the HUDF09 Team)

Focus Question 13-16

In what way are telescopes time machines?

formation that occurred in the early universe. The clustering of galaxies began relatively early in the life of the universe, too, as the most distant known cluster of galaxies is 12.6 Bly away.

The observational data now being accumulated from the period of time when the universe was less than a few billion years old are helping astronomers develop and test theories of the evolution of the early universe. We will explore this point further in Chapter 14.

We next turn our attention to active galaxies—energy sources of staggering power. While the short-lived outputs of supernovae boggle the mind, they represent miniscule amounts of energy compared to quasars, Seyfert galaxies, radio galaxies, and BL Lac objects. Quasars, for example, emit more energy each second than the Sun does in 200 years, and they continue to do so for millions of years. Evidence shows that most galaxies undergo periods of similar activity. Some truly remarkable activity must occur deep within quasars and other ultrahigh-energy sources to make them so luminous.

QUASARS

Our knowledge of quasars began in an amateur astronomer's backyard. Grote Reber built the first dedicated radio telescope in 1937 behind his home in Illinois, opening the realm of nonvisual astronomy. (Recall from Section 3-11 that in 1931 Karl Jansky discovered radio waves from sources in space using

an antenna originally designed for testing radio communication.) By 1944, Reber had detected strong radio emissions from sources in the constellations Sagittarius, Cassiopeia, and Cygnus. Two of these sources, Sagittarius A (Sgr A) and Cassiopeia A (Cas A), are in our Galaxy. The first is the galactic nucleus (see Section 13-4), and the second is a supernova remnant (see Chapter 12). However, Reber's third source, called Cygnus A (Cyg A), proved hard to categorize (Figure 13-53). Others quickly refined Reber's observations, but the mystery only deepened in 1954, when Walter Baade and Rudolph Minkowski, using the 200-in. optical telescope on Mount Palomar, discovered a strange-looking galaxy at the position of Cygnus A (Figure 13-53 inset).

The galaxy associated with Cyg A is very dim. Nevertheless, Baade and Minkowski managed to photograph its spectrum. They detected a redshift that corresponds to a speed of 14,000 km/s. According to the Hubble law, this speed indicates that Cyg A lies 635 million ly (194 Mpc) from Earth.

Because Cyg A is one of the brightest radio sources in the sky, its enormous distance intrigued astronomers. Although barely visible through the giant optical telescope at Palomar, Cyg A's radio waves can be picked up by amateur astronomers with backyard equipment. Its radio energy output must therefore be colossal. In fact, Cyg A shines with a radio luminosity 10^7 times as bright as the radio emission from an entire ordinary galaxy, such as Andromeda. The object that creates the Cyg A radio emissions has to be something extraordinary.

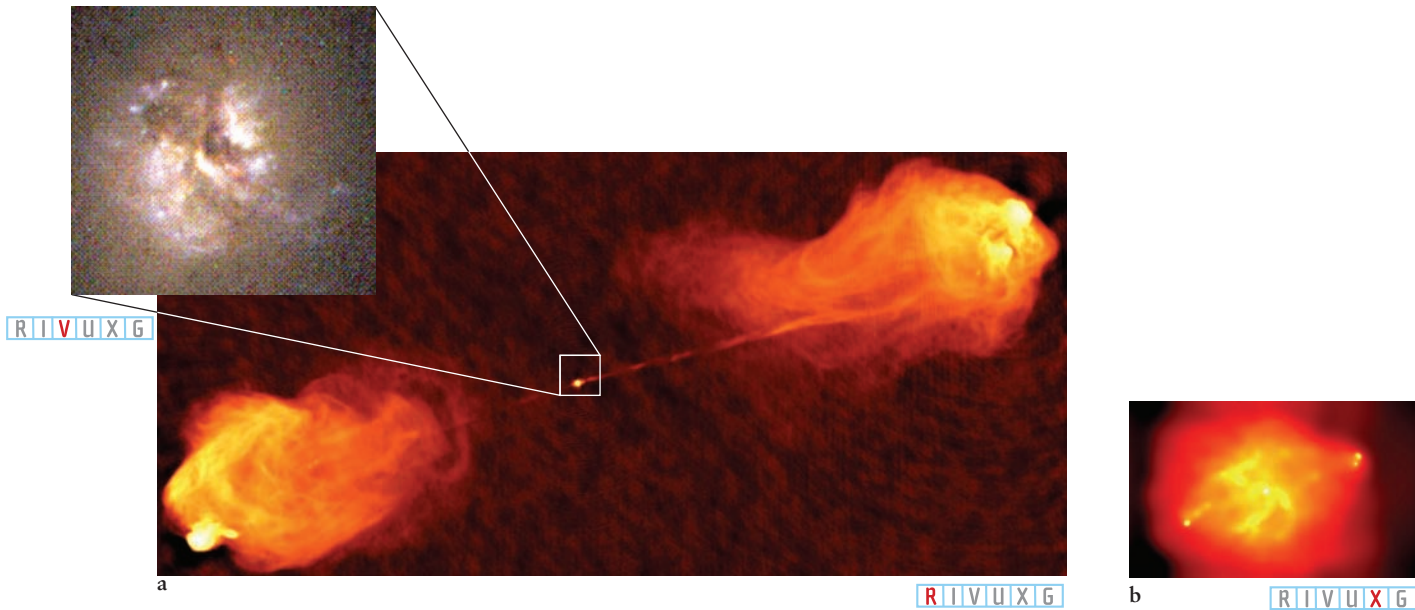


FIGURE 13-53 Cygnus A (3C 405) (a) Radio image produced from observations made at the Very Large Array. Most of the radio emissions from Cygnus A come from the radio lobes located on either side of the peculiar galaxy seen in the inset, a Hubble Space Telescope image. Each of the two radio lobes extend about 160,000 ly from the optical galaxy and contain a brilliant, condensed region of radio emission. **Inset:** At the heart of this system of gas lies a strange-looking galaxy that has a redshift that corresponds to a recessional speed of 5% of the speed of light. According to the Hubble

law, Cygnus A is therefore 635 million ly from Earth. Because Cygnus A is one of the brightest radio sources in the sky, this remote galaxy's energy output must be enormous. (b) An X-ray image of Cygnus A taken by the Chandra X-ray Observatory showing that some of the jet particles from the central source have spread throughout the galaxy and, striking other gas and dust, caused this debris to heat and emit X-rays. (a: NRAO/AUI/NSF; inset: William C. Keel, Robert Fosbury; b: NASA/UMD/A. Wilson et al.)

13-19 Quasars look like stars but have huge redshifts

Cygnus A is not the only powerful radio source in the far-distant sky. Starting in the 1950s, radio astronomers created long lists of radio sources. One of the most famous lists, the *Third Cambridge Catalogue*, was published in 1959. (The first two catalogues were filled with inaccuracies.) Even today, astronomers often refer to its 471 radio sources by their “3C numbers.” Cyg A, for example, is designated 3C 405, because it is the 405th source in the Cambridge list. Because of the extraordinary luminosity of Cyg A, astronomers were eager to learn whether any other sources in the 3C catalog had similar properties.

One interesting case was 3C 48. In 1960, Allan Sandage used the Palomar telescope to discover a “star” at the location of this radio source (Figure 13-54). Recall that stars are blackbodies whose peak intensity is typically visible light and whose radio emission is much less intense. Because ordinary stars are not strong sources of radio emission, 3C 48 had to be something unusual. Indeed, its spectrum showed a series of emission lines that, initially, no one could identify. Although 3C 48 was

clearly an oddball, many astronomers thought it was just another strange star in our Galaxy.

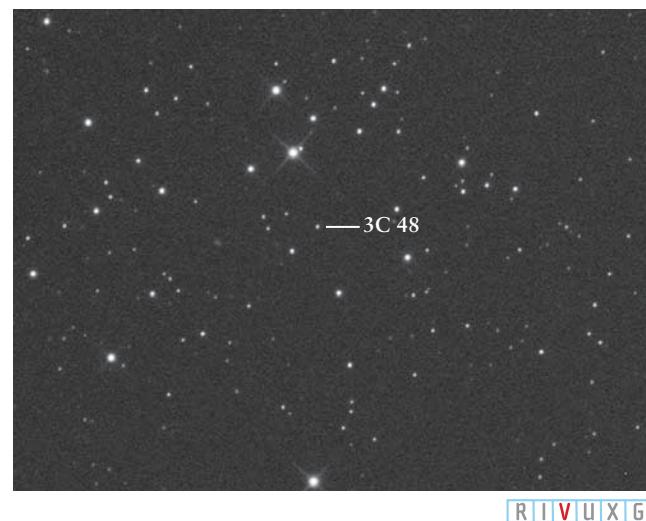


FIGURE 13-54 Quasar 3C 48 For several years, astronomers erroneously believed that this object was simply a peculiar, nearby star that happened to emit radio waves. Actually, the redshift of this starlike object is so great that, according to the Hubble law, it must be roughly 4 Bly away. (Anthony Ayiomamitis)

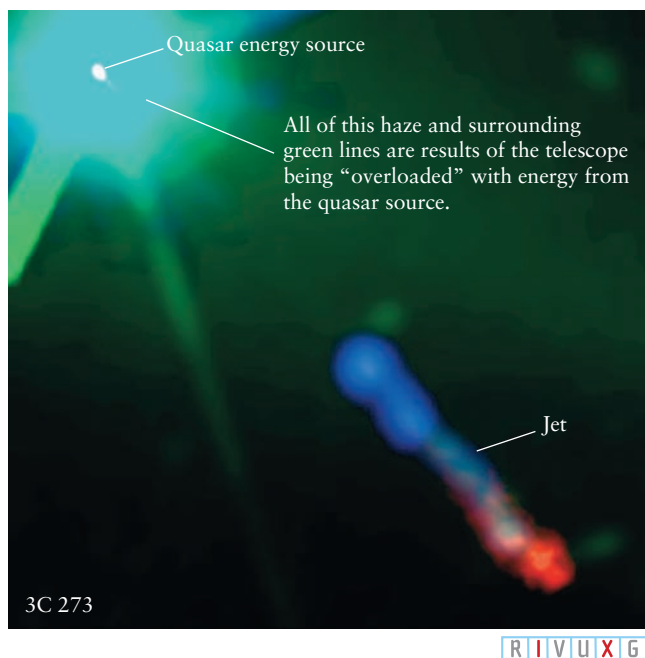


FIGURE 13-55 Quasar 3C 273 This combined X-ray and infrared view shows the starlike object associated with the radio source 3C 273 and the luminous jet it has created. The jet is also visible in the radio and visible parts of the spectrum. By 1963, astronomers determined that the redshift of this quasar is so great that, according to the Hubble law, it is nearly 2 Bly from Earth. (NASA/JPL-Caltech/Yale University/S. Jester, D. E. Harris, H. L. Marshall, K. Meisenheimer, H.-J. Roser, and R. Perley)

Another such “star,” called 3C 273, was discovered in 1962. Like 3C 48, this object (**Figure 13-55**) emits a series of bright spectral emission lines that no one could then identify. These spectral lines are brighter than the background radiation at other wavelengths, called the *continuum* (recall Kirchhoff’s laws in Section 3-14). **Figure 13-56** shows the spectrum of 3C 273, which looks nothing like a star’s blackbody spectrum with absorption lines. Recall from Chapter 3 that a spectrograph records energy intensity at different wavelengths. The emission lines appear as peaks, and absorption lines appear as valleys. The emission lines are caused by excited gas atoms that emit radiation at specific wavelengths.

A breakthrough finally came in 1963, when Maarten Schmidt at the California Institute of Technology identified four of the brightest spectral lines of 3C 273 as four spectral lines of hydrogen. However, these emission lines from 3C 273 are found at much longer wavelengths than the usual wavelengths of these lines. Schmidt concluded that the hydrogen lines are subjected to a substantial redshift (Doppler shift). Furthermore, the intense emission lines mean that something unusual is heating the gas.

Spectra for stars in our Galaxy exhibit comparatively small Doppler shifts, because these stars cannot move extremely fast relative to the Sun without soon escaping from the Galaxy. Schmidt thus concluded that 3C

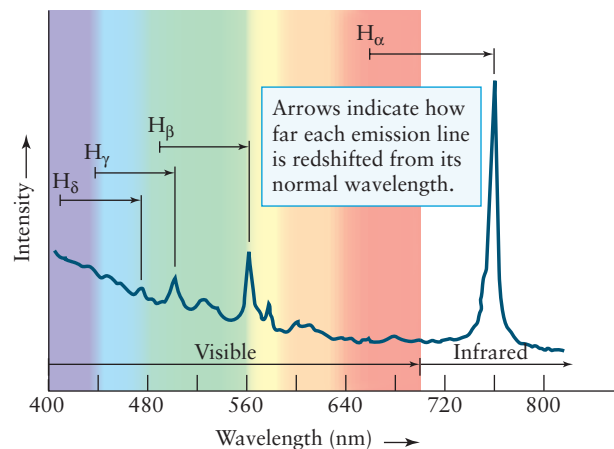


FIGURE 13-56 Spectra of 3C 273 The visible and infrared spectra of 3C 273 are dominated by four bright emission lines caused by hydrogen. This radiation is redshifted nearly 16% from its rest wavelengths.

273 is not a nearby star after all. Pursuing this conclusion, he promptly found that its redshift corresponds to a speed away from us of almost 16% the speed of light. According to the Hubble law, this huge redshift implies an impressive distance to 3C 273 of roughly 2 billion light-years (Bly).

Inspired by Schmidt’s success, astronomers looked again at the spectral lines of 3C 48. Sure enough, the emission lines were created by hydrogen, but with a redshift that corresponds to a velocity away from us of nearly one-third the speed of light. Therefore, 3C 48 must be nearly twice as far away as 3C 273, or about 4 Bly from Earth, assuming that the Hubble constant, H_0 , is 69 km/s/Mpc.

Figure 13-57 shows the spectrum of another object that is so far away that the expansion of the universe (see Section 13-17) gives it a redshift of 92% of the speed of light.

8 9 Because of their starlike appearances and strong radio emissions, 3C 48 and 3C 273 were dubbed **quasi-stellar radio sources**, a term soon shortened to **quasars**. Related objects, called **quasi-stellar objects (QSOs)**, also look like stars but have strong energy outputs in all different parts of the electromagnetic spectrum, with most being strongest in the infrared. More than 200,000 quasars and QSOs have been discovered. In this book, we will use the word *quasars* to cover QSOs, too.

All quasars have redshifts of at least 0.06, or 6%, of the speed of light. The closest known quasar, Cygnus A, is about 635 Mly from us. The most distant one observed to date has a redshift of 7.1, placing it 12.9 Bly away. It is moving away from Earth at greater than 99% of the speed of light. From the Hubble law, it follows that the distances to these high-redshift quasars are typically in the range of 10 Bly to more than 13 Bly. Plotting their distances, astronomers found that quasars began

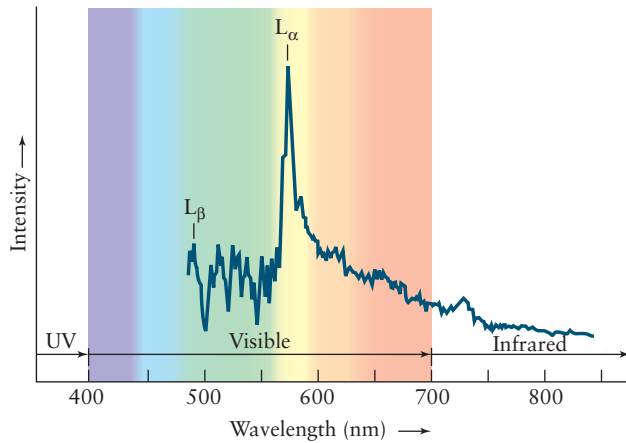


FIGURE 13-57 Spectrum of a High-Redshift Quasar The light from this quasar, known as PKS 2000-330, is so highly redshifted that spectral emission lines normally found in the far-ultraviolet (L_α and L_β) are seen at visible wavelengths. Note the many deep absorption lines on the short-wavelength side of L_α . These lines, collectively called the Lyman-alpha forest, are believed to be created by remote clouds of gas along our line of sight to the quasar. Hydrogen in these clouds absorbs photons from the quasar at wavelengths less redshifted than the quasar's L_α emission line.

forming early in the life of the universe (Figure 13-58), their number peaking about 2 billion years after it came into existence. The number of quasars dropped to near zero about 7 billion years ago. The reason for this evolution will become clear shortly.

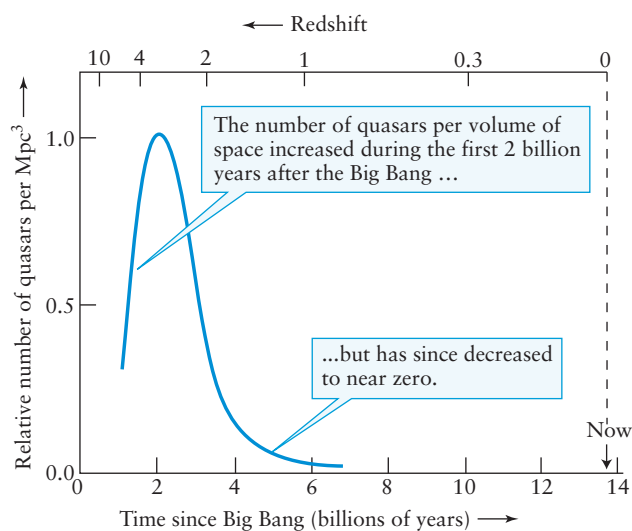


FIGURE 13-58 History of Quasar Formation The greater the redshift of a quasar, the farther it is from Earth and the further back in time we are seeing it. By observing the number of quasars found at different redshifts, astronomers can calculate how the density of quasars has changed over time.

A quasar emits a huge amount of energy from a small volume Galaxies are big and bright. A typical large galaxy, like our own Milky Way, contains hundreds of billions of stars and shines with the luminosity of 10 billion Suns. The largest and most luminous galaxies, the giant ellipticals, are only 10 times brighter. But beyond 8 Bly from Earth, even the brightest galaxies are too faint to be easily detected. Most ordinary galaxies are too dim to be detected at half that distance. If quasars can easily be seen more than 13 Bly away, they must be far more luminous than normal galaxies. Indeed, a typical quasar is 100 times brighter than our Milky Way.

In the mid-1960s, several astronomers discovered that quasars fluctuate in brightness. Going back over old images, they found that some newly identified quasars had actually been photographed in the past but had not been considered as anything special because they looked like stars. One photograph of 3C 273, for example, dates back to 1887. By carefully examining these old images, as well as recent ones, astronomers could see that *quasars occasionally flare up*. Prominent outbursts from quasar 3C 279 reached Earth around 1937 and 1943 (Figure 13-59). During these events, the luminosity of 3C 279 increased by a factor of about 25. At peak periods it was emitting at least 10,000 times as much energy as the entire Milky Way.

Fluctuations and their timescales Suppose that somewhere on a star or other spherical object an event occurred that began making the object more luminous and then returned it to normal brightness with a total variation or cycle time of, for example, 1 minute. If that fluctuation in luminosity were traveling at the speed of light, then it would take half a minute to cross the object and complete the process of brightening it, and another half a minute to reverse, so that the object would return

Focus Question 13-17

How are the spectra of quasars different from the spectra of stars?

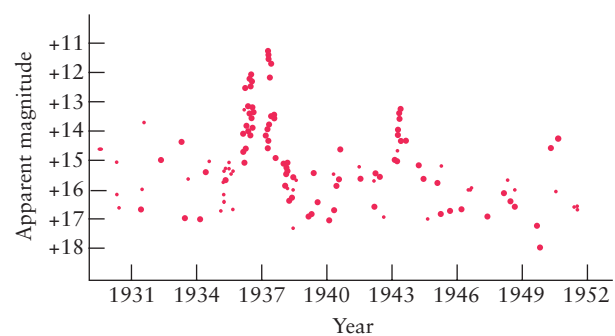


FIGURE 13-59 Brightness of 3C 279 This graph shows variations in the brightness of the quasar 3C 279. Note the especially large outburst observed in 1937. These data were obtained by carefully examining old photographic plates in the files of the Harvard College Observatory.

to its original brightness 1 minute after the event began. If the same fluctuation traveled more slowly, then the object would have to be smaller than half a light-minute across for it to return to its original state after 1 minute. Therefore, the object could not be more than half a light-minute across (ignoring the curve of a sphere). In this way, astronomers can calculate the maximum size an object can have from the timescale of its fluctuation.

Many quasars vary in brightness over only a few months, weeks, or days. In fact, X-ray observations reveal large variations in as little as 3 hours. This rapid flickering means that the source of the quasar's energy must be quite small by galactic standards. The energy-emitting region of a typical quasar—the “powerhouse” that blazes with the luminosity of 100 galaxies—is less than 1 light-day in diameter. Something must be producing the luminosity of 100 normal galaxies

from a volume with approximately the same diameter as our solar system! Before considering their energy source, let us examine other objects that will turn out to be powered by the same “engine,” as it is called by astronomers.

OTHER ACTIVE GALAXIES

Quasars are but one kind of very powerful energy source in space. The others include Seyfert galaxies, radio galaxies, double-radio sources, and BL Lacertae objects. Located at the centers of galaxies, these powerhouses are collectively called **active galactic nuclei**, or **AGN**. The galaxies that contain AGN are called **active galaxies**. Active galaxies have unusually bright, starlike (that is, a pinpoint as seen from Earth) nuclei; their spectra do not look anything like a star's black-body spectrum; they have strong emission lines; many are highly variable, meaning that their brightness levels change in different parts of the electromagnetic spectrum; they have jets of electrons and protons and beams of radiation emanating from their cores like the emissions observed for 3C 273; and most are more luminous than ordinary galaxies.

13-20 Active galaxies can be either spiral or elliptical

Seyfert galaxies Carl Seyfert at the Mount Wilson Observatory in California discovered the first active galaxies in 1943 among the spiral galaxies he was surveying. Now called **Seyfert galaxies**, these spirals reveal exceptionally bright, starlike nuclei and strong emission lines in their spectra. For example, the rich spectrum of NGC 4151 has many prominent emission lines. Some of these lines are produced by iron atoms with a dozen or more electrons stripped away, indicating that NGC 4151 contains some extremely hot gas. Individual Seyfert galaxies

also fluctuate in brightness. For example, at times the magnitude of NGC 4151 changes perceptibly over the course of a few days.

Another example of a Seyfert galaxy is NGC 1566 (Figure 13-60). At infrared wavelengths, this galaxy shines with the brilliance of 10^{11} Suns. This extraordinary luminosity has been observed to vary by as much as $7 \times 10^9 L_{\odot}$ over only a few weeks. In other words, the infrared power output of the nucleus of NGC 1566 increases and decreases by an amount nearly equal to the total luminosity of our entire Galaxy.

Many more Seyfert galaxies have been discovered in recent years. Approximately 10% of the most luminous galaxies in the sky are Seyferts. Some of the brightest Seyfert galaxies shine as brightly as faint quasars, with point sources in their centers. These are labeled Type 1 Seyferts. Seyfert galaxies whose centers are obscured by large amounts of gas and dust are Type 2 Seyferts.

Radio galaxies Some elliptical galaxies, called **radio galaxies** because of their strong radio emission, are like lower-energy quasars. Although their energy source is very small, like that of a quasar, the resulting radio emissions of these active galaxies are spread over very large areas, often hundreds of thousands of light-years across.

Some radio galaxies are accompanied by unusual optical features; namely, the galaxies look as though they are exploding, which they are not. Such galaxies are therefore called **peculiar galaxies** (denoted *pec*). *Peculiar*



FIGURE 13-60 Seyfert Galaxy NGC 1566 This Sc galaxy is a Seyfert galaxy some 50 Mly (16 Mpc) from Earth in the southern constellation Dorado (the Goldfish). The nucleus of this galaxy is a strong source of radiation whose spectrum shows emission lines of highly ionized atoms. (NASA/JPL-Caltech/R. Kennicutt [University of Arizona] and the SINGS Team)

Focus Question 13-18

Which can fluctuate faster—an object 1 million km in diameter or an object 1 billion km in diameter?

is not the classification of a galaxy shape, like *spiral*, *barred spiral*, or *elliptical*. Any of the Hubble classes of galaxies (see Sections 13-8 through 13-12) can be peculiar.

Figure 13-61 is a combined X-ray, optical, and radio image of the peculiar elliptical galaxy NGC 5128 (Centaurus A) in the southern constellation of Centaurus. The dark region that cuts the elliptical galaxy in two is actually a region of very thick dust through which little light from the galaxy behind it can penetrate. The galaxy is ejecting large volumes of hot X-ray-emitting gas, as well as jets of gas in two **radio lobes** shown in this figure. NGC 5128 is classified as an E0 (pec). The E0 denotes a circular elliptical galaxy, while (pec) means that this galaxy has the features of a peculiar galaxy.

Detailed observations of Centaurus A reveal that both its radio and X-ray jets (see Figure 13-61 inset) originate in the galaxy's nucleus. Particles and energy

stream out of the galaxy's nucleus toward the radio lobes at nearly the speed of light. Such galaxies, with two radio lobes, are now called **double-radio sources**. By 1970, radio astronomers had discovered dozens of other double-radio sources, usually in giant elliptical galaxies. Double-radio sources are among the brightest radio objects in the universe.

All double-radio sources appear to have a central engine that ejects electrons, protons, and magnetic fields outward along two oppositely directed jets. After traveling many thousands or even millions of light-years, this material slows down, allowing charged particles and magnetic fields to produce the radio radiation that we detect. This type of radio emission, called *synchrotron radiation* (see Section 13-4), occurs whenever energetic electrons, protons, or ions move in a spiral path within a magnetic field.

Focus Question 13-19

Is the Milky Way a peculiar galaxy? Why or why not?

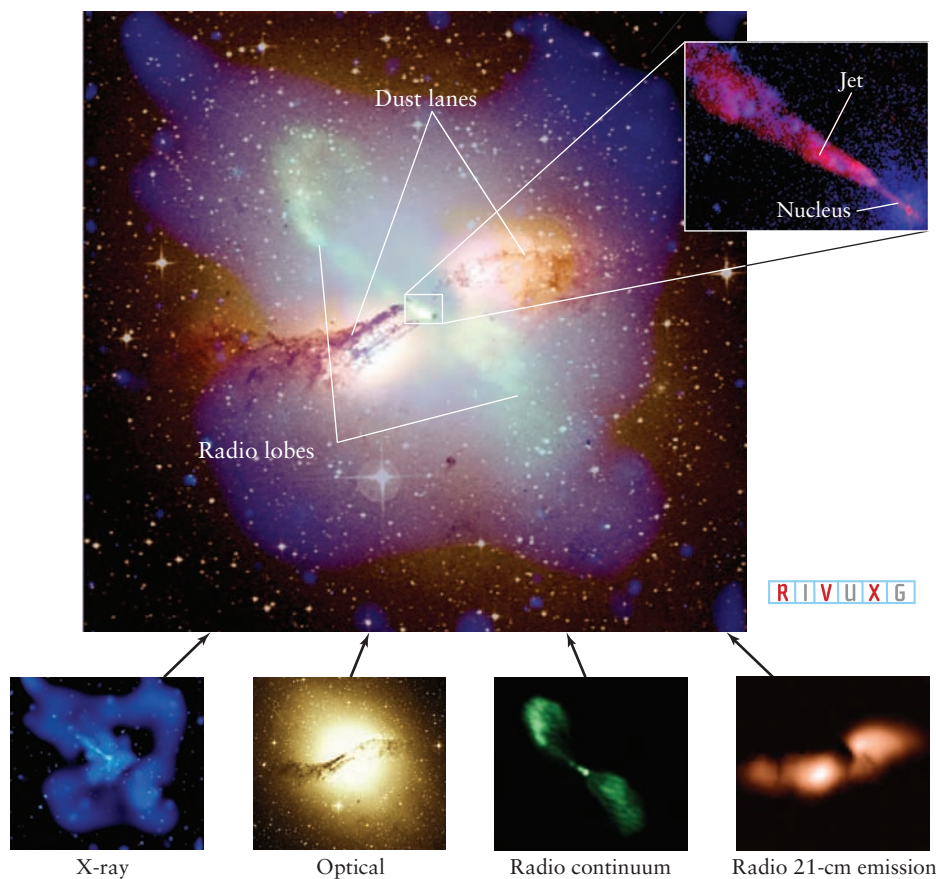


FIGURE 13-61 Peculiar Galaxy NGC 5128 (Centaurus A) This extraordinary radio galaxy is located in the constellation Centaurus, 11 million light-years from Earth. At visible wavelengths a dust lane crosses the face of the galaxy. Superimposed on this visible image is a false-color radio image (green) showing that vast quantities of radio radiation pour from matter ejected from the galaxy perpendicular to the dust lane, along with radio emission (rose-colored) along the dust lane, and X-ray emission (blue) detected by NASA's Chandra X-ray Obser-

vatory. The X-rays may be from material ejected by the black hole or from the collision of Centaurus A with a smaller galaxy. **Inset:** This X-ray image from the Einstein Observatory shows that NGC 5128 has a bright X-ray nucleus. An X-ray jet protrudes from the nucleus along a direction perpendicular to the galaxy's dust lane. (X-ray: NASA; radio 21-cm: NRAO/VLA/J. Van Gorkom/Schminovich et al.; radio continuum: NRAO/VLA/J. Conden et al.; optical: Digitized Sky Survey U.K. Schmidt Image/STScI; inset: X-ray: NASA)

The idea that a double-radio source emits jets of particles is supported by the existence of **head-tail sources**, so named because each such source appears to have a region of concentrated radio emission (the head) with two tails of gas that stream from it in opposite directions and then sweep back. NGC 1265, an active elliptical galaxy in the Perseus cluster of galaxies, shows these properties. NGC 1265 is known to be moving at a high speed (2500 km/s) relative to the Perseus cluster as a whole. In a radio map (Figure 13-62), its emission has a distinctly windswept appearance.

Protons and electrons ejected in the two jets from NGC 1265 are deflected by the galaxy’s passage through the sparse intergalactic medium. The head-tail source leaves behind a trail of particles, like the trail of smoke that pours from a rapidly moving steam train.

The giant elliptical galaxy M87 is a relatively weak radio galaxy, emitting only 1% as much energy as the Milky Way. Conversely, the distant, peculiar, elliptical galaxy Cygnus A (see Figure 13-53) is a powerful radio galaxy that emits 40 times as much energy as does our Galaxy. Pairs of orbiting supermassive black holes and accompanying double-radio sources have also been observed, such as 3C 75 (Figure 13-63). Table 13-2 summarizes the relative power emitted by the Sun, the Milky Way Galaxy, Seyfert galaxies, radio galaxies, and quasars.

BL Lacertae objects and blazars Although radio galaxies emit less energy from their nuclei than do quasars, there are also objects of the same size that from our vantage

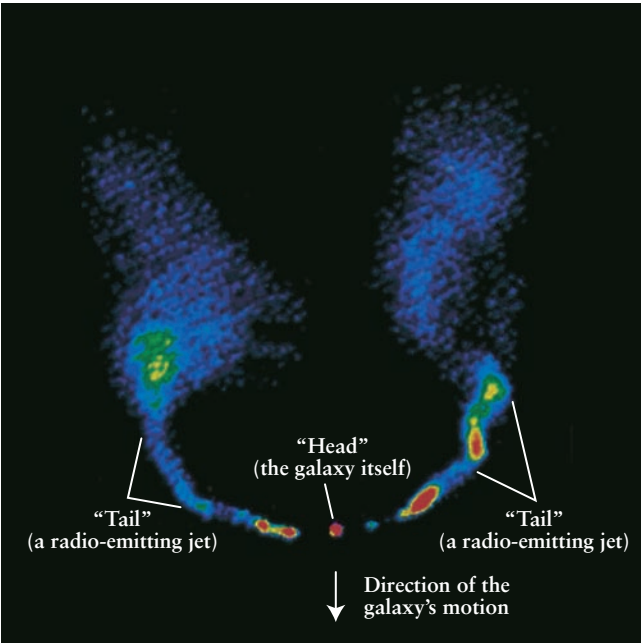


FIGURE 13-62 Head-Tail Source NGC 1265 This active elliptical galaxy is moving at a high speed through the intergalactic medium. Because of this motion, the two tail jets trail the galaxy at its head, giving this radio source a distinctly windswept appearance. (Image courtesy of NRAO/AUI and C. O’Dea & F. Owen)

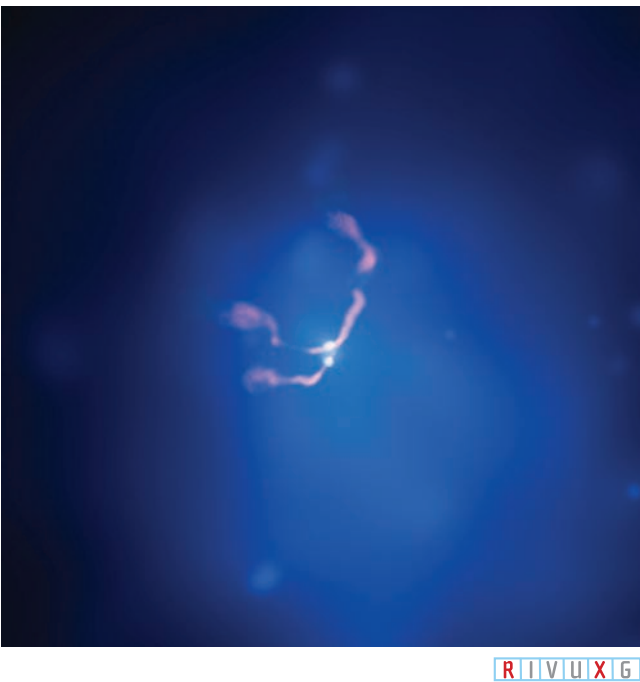


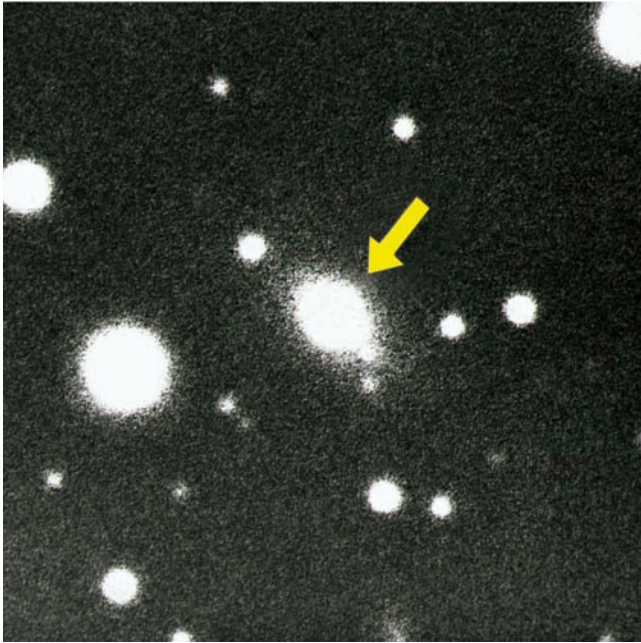
FIGURE 13-63 Binary Head-Tail Source This combined radio and X-ray image of 3C 75 shows the head-tail sources emanating from supermassive black holes in a pair of galaxies that are in the process of merging. The black holes are separated by 25,000 ly and are 300 million ly away from Earth. (NASA/CXC/D. Hudson, T. Reiprich et al. [AIFA])

point emit even more energy than quasars. These are the **BL Lacertae objects**. The name comes from their prototype, BL Lacertae (also called **BL Lac**) in the constellation Lacerta (the Lizard). BL Lac (Figure 13-64) was discovered in 1929, when it was mistaken for a variable star, largely because its brightness varies by a factor of 15 in only a few months. A BL Lac’s most intriguing characteristic is an unusually weak set of spectral lines. In other words, its spectrum is not as easily identified as those of stars or interstellar clouds. Nevertheless, the spectra enable astronomers to determine the redshifts of these objects.

BL Lac objects are believed to be elliptical galaxies with bright quasars at their centers, much as Seyfert galaxies are spiral galaxies with quasarlike centers. BL Lac objects change intensities—they are variable. Those that have periods of a day or less are also called **blazars**. In 2004, a blazar with 10 billion solar masses located 12.5 Bly away was dis-

TABLE 13-2 GALAXY AND QUASAR LUMINOSITIES

Object	Luminosity (watts)
Sun	4×10^{26}
Milky Way Galaxy	10^{37}
Seyfert galaxies	10^{36} – 10^{38}
Radio galaxies	10^{36} – 10^{38}
Quasar	10^{38} – 10^{42}



R I V U X G

FIGURE 13-64 BL Lacertae This photograph shows fuzz around BL Lacertae (arrow). The redshift of this fuzz indicates that BL Lacertae is about 900 Mly (280 Mpc) from Earth. BL Lac objects appear to be giant elliptical galaxies with bright quasarlike nuclei, much as Seyfert galaxies are spiral galaxies with quasarlike nuclei. BL Lac objects contain much less gas and dust than do Seyfert galaxies. (T. D. Kinman, NOAO/AURA/NSF)

covered. It is intriguing that this incredibly massive object formed only a billion years after the universe formed.

SUPERMASSIVE ENGINES

Quasars, BL Lac objects, Seyfert galaxies, and radio galaxies are excellent examples of how astronomers can use

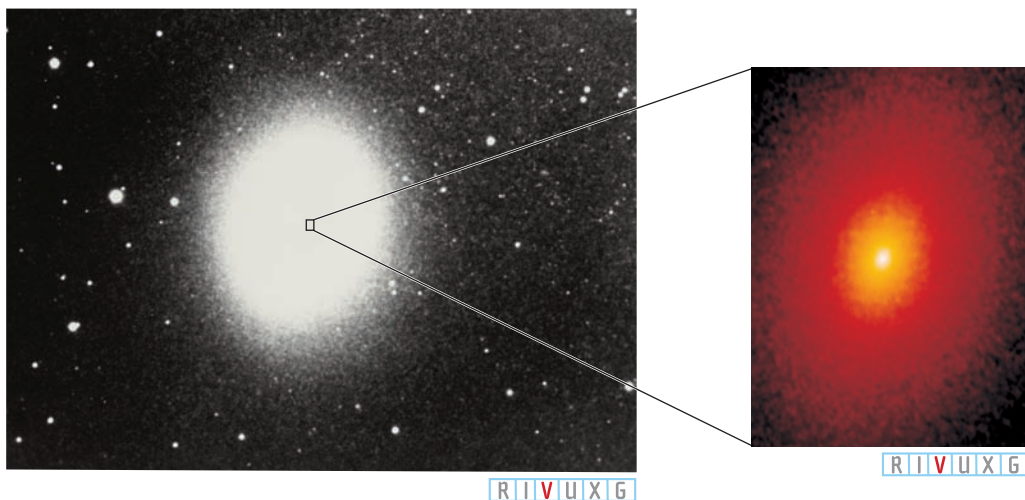
one mechanism to explain an apparently wide variety of objects in space. In 1968, the British astronomer Donald Lynden-Bell suggested that the gravitational field of a supermassive black hole (see Chapter 12) could cause the enormous releases of energy for quasars and other active galaxies.

13-21 Supermassive black holes exist at the centers of most galaxies

As we saw in Chapter 12, black holes with a wide range of masses are being discovered throughout the universe. Supermassive black holes (those exceeding a million solar masses or so) have been detected in a rapidly growing number of galaxies, including our Milky Way. To emphasize the similarities in active galaxies, we consider five further examples: M31, M32, M87, M104, and NGC 3115.

The Andromeda Galaxy (M31) is the second largest, second most massive galaxy in the Local Group (after the Milky Way). In the mid-1980s, several astronomers made careful spectroscopic observations of the core of M31. Using measurements of Doppler shifts, they determined that stars within 50 ly of the galactic core are orbiting the nucleus at exceptionally high speeds, which suggests that a massive object is located at the galaxy's center. Without the gravity of such an object to keep the stars in their high-speed orbits, they would have escaped from the core region long ago. From such observations, astronomers estimate the mass of the central object in M31 to be about 50 million solar masses. That much matter confined to such a small volume strongly suggests that a supermassive black hole is located there.

Near M31 is a small elliptical galaxy, called M32 (**Figure 13-65**). High-resolution spectroscopy indicates that stars close to the center of M32 are also orbiting this galaxy's nucleus at unusually high speeds, which is



R I V U X G

R I V U X G

FIGURE 13-65 Elliptical Galaxy M32 This small galaxy is a satellite of M31, a portion of which is seen at the left in this wide-angle photograph. Both galaxies are roughly 2.5 million ly from Earth. **Inset:** High-resolution image from the Hubble

Space Telescope that shows the center of M32. Note the concentration of stars at the nucleus of the galaxy. The nucleus is only 175 ly across. (NOAO/Science Photo Library/Science Source; inset: Tod R. Lauer/NASA/ESA)

explained by the presence of a supermassive black hole there, too. A picture taken by the Hubble Space Telescope (Figure 13-65 inset) shows that the concentration of stars at the core of M32 is truly remarkable. Its number density of stars is more than 100 million times greater than the density of stars in the Sun's neighborhood. The concentration of stars and their high

Focus Question 13-20

What makes it so hard for us to see deep into elliptical galaxies?

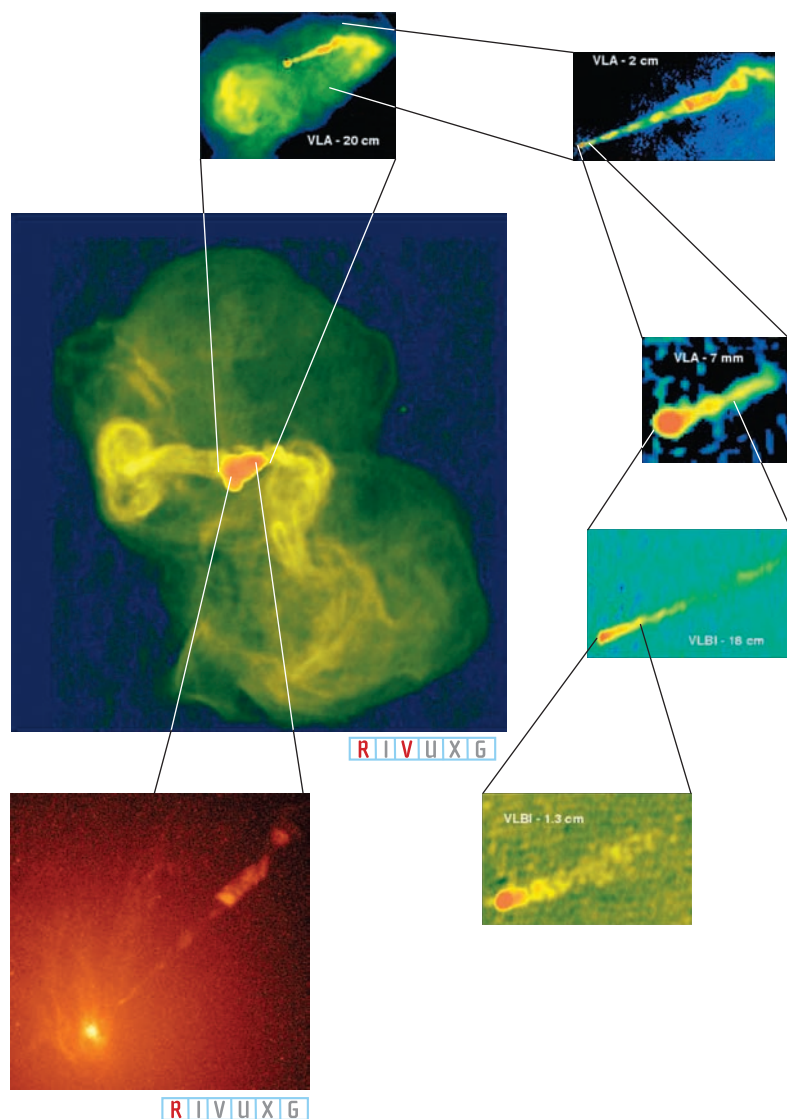


FIGURE 13-66 Giant Elliptical Galaxy M87 M87 is located near the center of the sprawling, rich Virgo cluster, which is about 50 million ly from Earth. Embedded in this radio image of gas is the galaxy M87 from which the gas has been ejected (**bottom inset**). Images at different radio and visible wavelengths reveal a variety of details about the structure of the jets of gas. M87's extraordinarily bright nucleus and the gas jets result from a 3-billion- $M_{\odot}$ black hole, whose gravity causes huge amounts of gas and an enormous number of stars to crowd around it. (Center: Image courtesy of NRAO/AUI/NSF; bottom: Holland Ford, Space Telescope Science Institute/Johns Hopkins University; center image insets: Richard Harms, Applied Research Corp.; Zlatan Tsvetanov, Arthur Davidsen, and Gerard Kriss at Johns Hopkins; Ralph Bohlin and George Hartig at Space Telescope)

speeds strongly support the belief that a supermassive black hole also exists at the center of M32.

One early candidate galaxy for study by the Hubble Space Telescope was the giant elliptical galaxy M87 (**Figure 13-66**). Located some 50 Mly from Earth, M87 is an active galaxy that has long been recognized as unusual. In 1918, Heber Curtis at the Lick Observatory reported that the center of M87 has “a curious straight ray . . . apparently connected with the nucleus by a thin line of matter.” Figure 13-66 shows radio and visible images of M87 to scale. M87 is also a powerful source of X-rays. X-rays emitted by such small sources are a signature of black holes (as discussed in Section 12-14).

The Hubble Space Telescope image of M87 in Figure 13-66 shows an exceptionally bright, starlike nucleus and a surrounding disk of gas with trailing spiral arms. To produce this fiery glow, stars must be packed so tightly at the center of M87 that their density is at least 300 times greater than that normally found at the centers of giant ellipticals. The motions of the stars in this central clustering support the hypothesis that a black hole with nearly 3 billion $M_{\odot}$ resides at the center of M87.

Even edge-on spiral galaxies are revealing their innermost secrets. John Kormendy used a 3.6-m telescope on Mauna Kea to examine the core of M104 (**Figure 13-67**) spectroscopically. Once again, high-

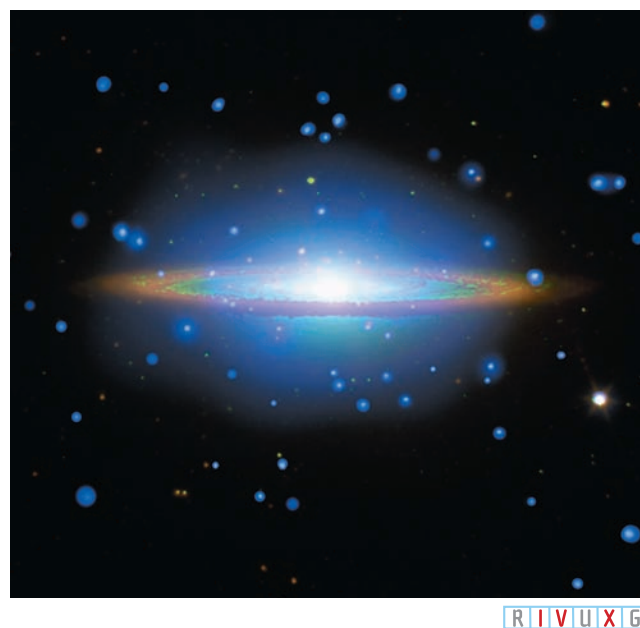


FIGURE 13-67 Sombrero Galaxy (M104) This spiral galaxy in Virgo is nearly edge-on to our Earth-based view. Spectroscopic observations indicate that a billion-solar-mass black hole is located at the galaxy's center. You can see the bright region in the galaxy's center created by stars and gas that orbit the black hole. (infrared: NASA/JPL-Caltech/Univ. AZ/R. Kennicutt/SINGS Team; Optical: NASA/STScI/AURA/Hubble Heritage; X-ray: NASA/UMass/Q. D. Wang et al.)

speed gas orbiting the galaxy's nucleus was found. These observations suggest that the center of this galaxy is dominated by a supermassive black hole that contains a billion solar masses. Similar spectroscopic observations of the edge-on disk galaxy NGC 3115 also reveal evidence of a billion-solar-mass black hole at its core.

13-22 Jets of protons and electrons ejected from around black holes help explain active galaxies

10 The gravitational energy associated with supermassive black holes at the centers of galaxies creates huge jets of gas comprised of protons and electrons. To see how this works, consider a black hole at the center of a young galaxy filled with gas and dust. Because the centers of galaxies are congested places, a black hole located there will capture a massive accretion disk of this debris (Figure 13-68a). If the black hole is not rotating (a Schwarzschild black hole) or if it is a slowly rotating Kerr black hole, then most of the mass approaching it from the accretion disk will spiral down into it. This does not last for long, because as this debris spirals in, the black hole gains angular momentum from the in-spiraling debris and therefore begins rotating faster and faster. By the time the black hole has doubled its original mass, its event horizon is whipping around at nearly the speed of light. The gas and dust spiraling toward a rapidly rotating Kerr black hole has much more difficulty entering the black hole because the motion of the black hole literally drags the space and the debris surrounding it around and around (see Section 12-12), thereby slowing the infall. The formation of jets occurs in this debris.

According to Kepler's third law, the material in the inner regions of such a disk orbits the most rapidly. The inner matter, therefore, constantly interacts with the more slowly moving gases in the outer regions, and the friction between them heats all of the gas. As energized gases spiral toward the black hole, they are compressed and heated to millions of degrees. In the vicinity of a rapidly rotating Kerr black hole, this temperature creates great pressure, causing some of the hot gas to expand and eventually squirt out where the resistance to its expansion is lowest—namely, perpendicular to the accretion disk. These dynamics are analogous to the accretion disk and jets that occur around solar-mass black holes, as discussed in Section 12-16.

In 1995, the Hubble Space Telescope took a picture of the giant elliptical galaxy NGC 4261 (Figure 13-68b) that astronomers interpret as showing this set of events. A disk of gas and dust about 800 ly in diameter is seen orbiting a supermassive black hole. The speed of the material in this disk indicates that it is held in orbit by a 1.2-billion- $M_{\odot}$ object. The inset in Figure 13-68b is a Hubble image of the galaxy's nucleus. The dark, horizontal oval in the picture is our oblique view of the thick disk of gas and dust surrounding the black hole. The central region of this disk

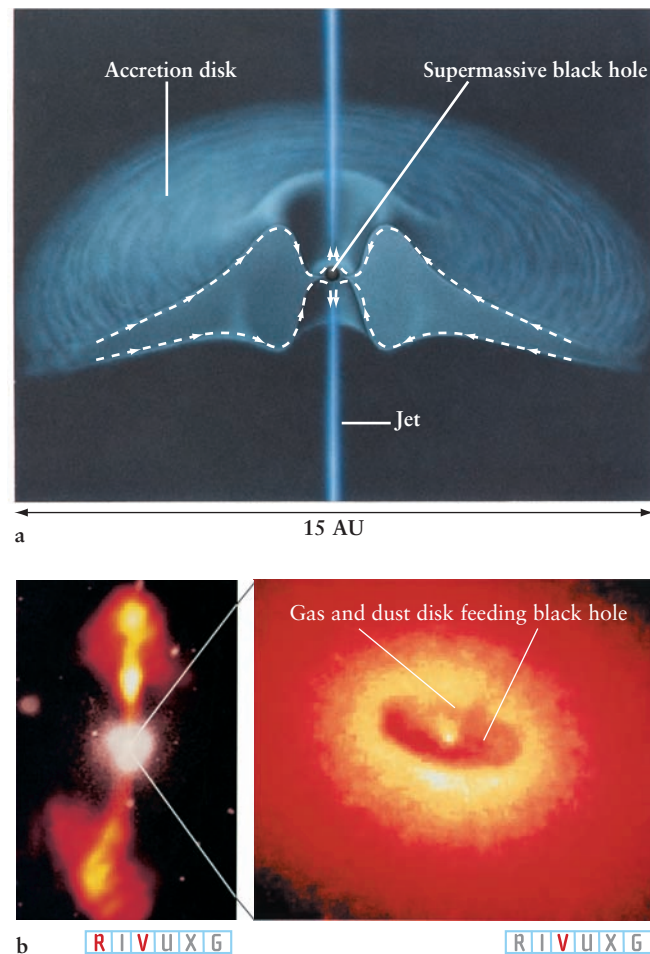


FIGURE 13-68 Supermassive Black Holes as Engines for Galactic Activity (a) Analogous to the activity depicted in Figure 12-46, in the accretion disk around a supermassive black hole, inswirling gas heats and expands. Pulled inward, compressed, and heated further, some of it is eventually expelled perpendicular to the disk in two jets. (b) The giant elliptical galaxy NGC 4261 is a double-radio source located in the Virgo cluster, about 100 million ly from Earth. An optical photograph of the galaxy (white) is combined with a radio image (orange and yellow) to show both the visible galaxy, which does not emit much radio energy, and its jets, which do. **Inset:** This Hubble Space Telescope image of the nucleus of NGC 4261 shows a disk of gas and dust about 800 ly (250 pc) in diameter, orbiting a supermassive black hole. (Science Source)

is the black hole's accretion disk. Ground-based radio and optical observations show a double-lobed structure (Figure 13-68b) that is bisected by the oval.

What confines the ejected gas to narrow jets? At first, the gas still swirling inward toward the black hole prevents the jets from spreading. To see why, consider the analogy of water spraying straight up out of the nozzle of a hose on Earth. Pulled down by gravity, the stream of water broadens and fans out through a wide angle in



FIGURE 13-69 Focusing Jets by Pressure (a) If a high-speed jet of gas or liquid encounters little pressure (from the surrounding air, in this image), then it will spread out. (b) If the jet encounters high pressure, such as occurs when it enters water, then it will maintain longer its shape as a column. (a: Repina Valeriya/Shutterstock; b: 2004 Richard Megna—Fundamental Photographs)

the air (Figure 13-69a). If the nozzle is placed in a swimming pool, however, a stream of liquid will not fan out as much at first (Figure 13-69b). The swirling mass of the accretion disk plays the role of the water, keeping the ejected gases from spreading out.

Beyond the disk, magnetic fields come into play. Because it is so hot, this gas is ionized and the inward spiral of these ions generates a magnetic field (see Section 6-5). As you can see in Figure 13-70, the magnetic fields encircle and guide the ions and electrons rushing out of the vicinity of the black hole.

After traveling for hundreds of thousands of years or more, the jets of gas interact with enough preexisting interstellar and intergalactic gas to be spread out and eventually stopped, thereby forming the lobes and other features (see Figure 13-53a). In 2003, astronomers using the Chandra X-ray Observatory observed gas jetting out from a supermassive black hole in a galaxy at the center of the Perseus cluster of galaxies and striking gas that resides between galaxies. This *intracluster* gas became compressed by the jet and generated concentric rings or shells of sound waves 35,000 ly in wavelength. Put another way, the sound created by the jet is a B-flat that is 57 octaves below middle C.

This “rocket nozzle” model explains not only double-radio sources—one coming out each side of the accretion disk—but also the jets and beams we see protruding from other active galaxies. We can now tie together all of the various active galaxies discussed in this chapter. *The major difference between the various types of active galaxies is the angle at which we view the central engine.* As Figure 13-71 shows, an observer sees a radio galaxy or a double-radio source when the accretion disk is viewed

nearly edge-on because the jets are nearly in the plane of the sky. At a steeper angle, a quasar is seen. If one of the jets is aimed almost directly at Earth, the galaxy is a BL Lac object. The Seyferts, active spiral galaxies, are distinguished by whether we can see the bright core (Type 1) or not (Type 2). Their placements in Figure 13-71 are still under investigation.

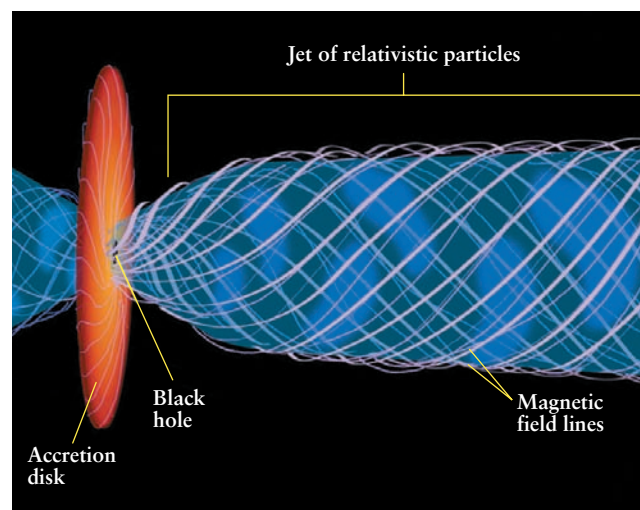


FIGURE 13-70 Focusing Jets by Magnetic Fields The hot, ionized accretion disk spirals around the central black hole. This motion creates a magnetic field that is twisted into spring-shaped spirals on both sides of the disk. Some of the accretion disk’s gas falling toward the black hole is overheated and squirted at high speeds into the two tubes created by the magnetic fields. The fields keep the gas traveling directly outward from above and below the disk, thus creating the two jets.

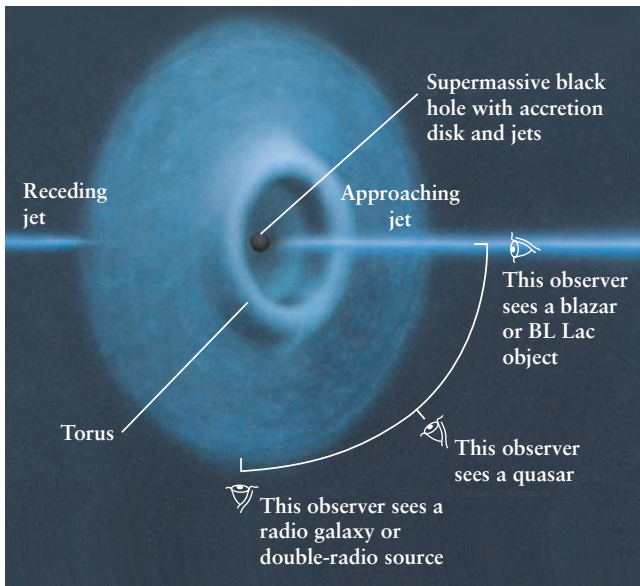


FIGURE 13-71 Orientation of the Central Engine and Its Jets BL Lacertae objects, quasars, and double-radio sources appear to be the same type of object viewed from different directions. If one of the jets is aimed almost directly at Earth, we see a BL Lac object. If the jet is somewhat tilted to our line of sight, we see a quasar. If it is tilted farther, we see an active galaxy. If the jets are nearly perpendicular to our line of sight, we see a double-radio source. The central region of the system is shown in Figures 13-68a and 13-70.

There are many active galaxies that we *should* see but do not. These objects have such thick and dusty accretion disks that little visible light or X-rays get through them. We cannot see the energy sources, for example, of quasars using visible light or X-ray telescopes in such situations. However, infrared radiation should pass through the gas and dust. Thus, in 2005, astronomers used the Spitzer Space Telescope to study regions of galaxies in which quasars should exist. In just one tiny patch of sky, they found 21 quasars. Extrapolating to the total volume of space that contains quasars, they expect “stealth” active galaxies to exist by the millions or more.

The other major factor in the observed properties of active galaxies and their evolutionary history is the amount of material available to make the accretion disk. Overlap in the brightness levels of different types of active galaxies occurs because central black holes and their accretion disks have different masses. The intensities of active galaxy emissions change as the amount of gas and dust in their accretion disks varies.

The amount of gas available in a galactic nucleus also changes with time. When supermassive black holes are young, they are surrounded by vast amounts of gas, which create massive accretion disks. Over time, much of this gas either enters the black hole or jets away from the disk, as just discussed. When an accretion disk is de-

Insight Into Science

Occam’s Razor Revisited It is possible to create *separate* theories for quasars, BL Lac objects, and other active galaxies. However, scientists prefer a comprehensive single theory that explains as many things as possible, because it incorporates common properties of the various phenomena. This search for simple, more powerful explanations is the central engine of progress in science.

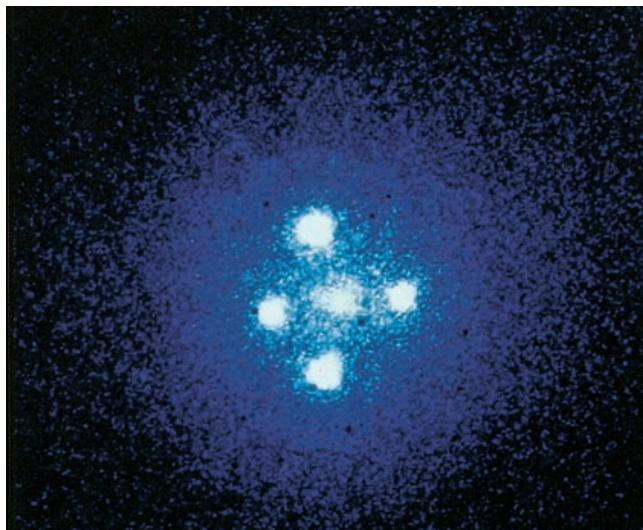
pleted, that active galaxy becomes more and more quiet. Because the galaxies we observe close to us are about the same age as the Milky Way, over 13 billion years old, their central black holes have already either consumed or ejected most of the gas and dust in their neighborhoods. Therefore, most have small or no accretion disks around the massive black holes in their nuclei and, hence, they are no longer AGNs. Thus, we do not see many active galaxies close to us. However, when galaxies collide, gas can be available to enable quiet galaxies to once again become active.

Interestingly, the presence of an AGN causes the star formation rates in a galaxy to decrease, compared to the activity in the same galaxy without an AGN. The reason for this is that star formation occurs when gas and dust become Jeans’ unstable (see Section 4-2), a phenomenon that is sensitive to the temperature of the gas and dust. The cooler the debris is, the more susceptible it is to collapse and star formation. AGNs heat the gas in their galaxies, making it less able to collapse and form new stars. When an AGN turns off, the gas and dust throughout the galaxy cool and the star formation rate increases.

13-23 Gravity focuses light from quasars

The radiation emitted by all quasars is subject to the gravitational lensing described by Einstein’s theory of general relativity (see Sections 12-11 and 13-7). Like the light from stars behind the Sun, the light from quasars is deflected as it travels past other galaxies or intergalactic gas and dark matter toward us. As we saw in Section 13-16 this gravitational lensing can lead to our receiving several images of objects. **Figure 13-72** shows four images of a quasar in a configuration called an **Einstein cross**. See also Figure 13-48c. If the background object is exactly behind an intervening quasar, the light is actually focused as a ring, called an **Einstein ring** (see Figure 13-48b), rather than as several separate images.

As you know, light travels at a finite speed. For example, light from the Sun takes about 8 minutes to reach Earth. We see the Sun as it was 8 minutes ago. As noted earlier in this chapter, the farther an object is from us, the longer it takes light and other radiation from it to reach



R I V U X I G

FIGURE 13-72 Gravitational Lensing of Quasars Image from the Hubble Space Telescope that shows the gravitational lensing of a quasar in the constellation of Pegasus. The quasar, about 8 billion ly from Earth, is seen as four separate images that surround a galaxy that is only 400 million ly away. This pattern is called an Einstein cross. The diffuse image at the center of the Einstein cross is the core of the intervening galaxy. The physical effect that creates these multiple images is the same as that seen for galaxies, as depicted in Figure 13-48. (NASA, ESA, and STScI)

Focus Question 13-21

What other astronomical objects have been observed after their light passed through a gravitational lens?

Earth. We see a galaxy 1 Bly away as it was a billion years ago. When we look at quasars, we are seeing objects as they appeared when the universe was younger still. In the last few years, astronomers using the Sloan Digital Sky Survey have discovered quasars at a

distance corresponding to when the universe was only 770 million years old.

SUMMARY OF KEY IDEAS

The Milky Way

- A century ago, astronomers were divided on whether or not the Milky Way Galaxy and the universe were the same thing.
- The inconclusive Shapley–Curtis debate was the first major public discussion between astronomers as to whether the Milky Way contains all the stars in the universe.
- Cepheid variable stars are important in determining the distance to other galaxies.
- Edwin Hubble proved that there are other galaxies far outside of the Milky Way.

The Structure of Our Galaxy

- Our Galaxy has a disk about 100,000 ly in diameter and about 2000 ly thick, with a high concentration of interstellar dust and gas. It contains around 200 billion stars.
- Interstellar dust obscures our view into the plane of the galactic disk at visual wavelengths. However, hydrogen clouds can be detected beyond this dust by the 21-cm radio waves emitted by changes in the relative spins of electrons and protons in the clouds, as well as by other nonvisible emissions.
- The center, or galactic nucleus, has been studied at gamma-ray, X-ray, infrared, and radio wavelengths, which pass readily through intervening interstellar dust and H II regions that illuminate the spiral arms. These observations have revealed the dynamic nature of the galactic nucleus, but much about it remains unexplained.
- A supermassive black hole of about $4.3 \times 10^6 M_{\odot}$ exists in the galactic nucleus.
- The galactic nucleus of the Milky Way is surrounded by a flattened sphere of stars, called the central bulge, through which a bar of stars and gas extends.
- A disk with two bright arms of stars, gas, and dust spirals out from the ends of the bar in the galactic central bulge.
- Young OB associations, H II regions, and molecular clouds in the galactic disk outline huge spiral arms where stars are forming.
- The Sun is located about 26,000 ly from the galactic nucleus, between two spiral arms. The Sun moves in its orbit at a speed of about 878,000 km/h (550,000 mi/h) and takes about 230 million years to complete one orbit around the center of the Galaxy.
- The entire Galaxy is surrounded by two halos of matter. The inner halo includes a spherical distribution of globular clusters and field stars, as well as large amounts of dark matter. It orbits in the same general direction as the disk. The outer halo is composed of dark matter and very old stars, which have retrograde orbits.

Types of Galaxies

- The Hubble classification system groups galaxies by their shapes into five major types: spiral, barred spiral, lenticular, elliptical, and irregular.
- The arms of spiral and barred spiral galaxies are sites of active star formation.
- According to the theory of self-propagating star formation, spiral arms of flocculent galaxies are caused by the sequential births and deaths of stars over extended regions of a galaxy. Differential rotation of a galaxy stretches these star-forming regions into elongated arches of stars and nebulae that we see as spiral arms.

- According to the spiral density wave theory, spiral arms of grand-design galaxies are caused by density waves. The gravitational field of a spiral density wave compresses the interstellar clouds that pass through it, thereby triggering the formation of stars, including OB associations, which highlight the arms.
- Elliptical galaxies contain much less interstellar gas and dust than do spiral galaxies; little star formation occurs in elliptical galaxies.
- Irregular galaxies are rich in gas and dust, and star formation occurs in them.
- Lenticular galaxies are disk galaxies without spiral arms.

Clusters and Superclusters

- Galaxies group into clusters rather than being randomly scattered through the universe.
- A rich cluster contains at least a thousand galaxies; a poor cluster may contain only a few dozen up to a thousand galaxies. A regular cluster has a nearly spherical shape with a central concentration of galaxies; in an irregular cluster, the distribution of galaxies is asymmetrical.
- Our Galaxy is a member of a poor, irregular cluster, called the Local Group.
- Rich, regular clusters contain mostly elliptical and lenticular galaxies; irregular clusters contain more spiral and irregular galaxies. Giant elliptical galaxies are often found near the centers of rich clusters.
- Each galaxy is held together with the aid of dark matter.
- No cluster of galaxies has an observable mass large enough to account for the observed motions of its galaxies; a large amount of dark matter must be present between the galaxies.
- Hot intergalactic gases emit X-rays in rich clusters.
- When two galaxies collide, their stars initially pass each other, but their interstellar gas and dust collide violently, either causing gas and dust to be stripped from the galaxies or triggering prolific star formation. The gravitational effects of a galactic collision can cast stars out of their galaxies into intergalactic space.
- Galactic mergers occur. A large galaxy in a rich cluster may also grow steadily through galactic cannibalism.

Superclusters in Motion

- A simple linear relationship exists between the distance from Earth to galaxies in other superclusters and the redshifts of those galaxies (a measure of the speed at which they are receding from us). This relationship is the Hubble law: Recessional velocity = $H_0 \times \text{distance}$, where H_0 is the Hubble constant.

- Astronomers use standard candles—Cepheid variables, the brightest supergiants, globular clusters, H II regions, supernovae in a galaxy, and the Tully–Fisher relation—to calculate intergalactic distances. Because of difficulties in measuring the distances to remote galaxies, the value of the Hubble constant, H_0 , is not known with complete certainty.
- The development of radio astronomy in the late 1940s led to the discovery of very powerful and extremely distant energy sources.

Quasars and Other Active Galaxies

- An active galaxy is an extremely luminous galaxy that has one or more unusual features: an unusually bright, starlike nucleus; strong emission lines in its spectrum; rapid variations in luminosity; and jets or beams of radiation that emanate from its core. Active galaxies include quasars, Seyfert galaxies, radio galaxies, double-radio sources, and BL Lacertae objects.
- A quasar, or quasi-stellar radio source, is an object that looks like a star but has a huge redshift. This redshift corresponds to a distance of billions of light-years from Earth, according to the Hubble law.
- To be seen from Earth, a quasar must be very luminous, typically about 100 times brighter than an ordinary galaxy. Relatively rapid fluctuations in the brightness levels of some quasars indicate that they cannot be much larger than the diameter of our solar system.
- An active spiral galaxy with a bright, starlike nucleus and strong emission lines in its spectrum is categorized as a Seyfert galaxy.
- An active elliptical galaxy is called a radio galaxy. It has a bright nucleus and a pair of radio-bright jets that stream out in opposite directions.
- BL Lacertae (BL Lac) objects (some of which are called blazars) have bright nuclei whose cores show relatively rapid variations in luminosity.
- Double-radio sources contain active galactic nuclei located between two characteristic radio lobes. A head-tail radio source shows evidence of jets of high-speed particles that emerge from an active galaxy.

Supermassive Engines

- Many galaxies contain huge concentrations of matter at their centers.
- Some matter that spirals in toward a supermassive black hole is squeezed into two oppositely directed beams that carry particles and energy into intergalactic space.
- The energy sources from quasars, Seyfert galaxies, BL Lac objects, radio galaxies, and double-radio sources are believed to be matter ejected from the accretion disks that surround supermassive black holes at the centers of galaxies.

WHAT DID YOU THINK?

1. *What is the shape of the Milky Way Galaxy?* The Milky Way is a barred spiral galaxy. A bar of stars, gas, and dust runs through its central region. It has two major spiral arms, several minor arms, and is surrounded by a complex spherical halo system of stars and dark matter.
2. *Where is our solar system located in the Milky Way Galaxy?* The solar system is between the Sagittarius and Perseus spiral arms, about 26,000 ly from the center of the Galaxy (about halfway out to the visible edge of the galactic disk).
3. *Is the Sun moving through the Milky Way Galaxy and, if so, about how fast?* Yes. The Sun orbits the center of the Milky Way Galaxy at a speed of 878,000 km/h (550,000 mi/h).
4. *Are most of the stars in spiral galaxies located in their spiral arms?* No. The spiral arms contain only 5% more stars than the regions between the arms.
5. *Do all galaxies have spiral arms?* No. Galaxies may be spiral, barred spiral, lenticular, elliptical, or irregular. Only spirals and barred spirals have arms.
6. *Are galaxies isolated objects?* No. Galaxies are grouped in clusters, and clusters are grouped in superclusters.
7. *Is the universe contracting, unchanging in size, or expanding?* The universe is expanding.
8. *What does “quasar” stand for?* The term *quasar* stands for quasi-stellar radio source.
9. *What do quasars look like?* They look like stars, but they emit much more energy than any star.
10. *Where do quasars get their energy?* A quasar is powered by a supermassive black hole with millions or billions of solar masses at the center of a galaxy.

Review Questions

1. What is located in the nucleus of the Galaxy?
 - a. a globular cluster
 - b. a spiral arm
 - c. a black hole
 - d. the solar system
 - e. a MACHO
2. Which statement about the Milky Way Galaxy is correct?
 - a. Our Galaxy is but one of many galaxies.
 - b. Our Galaxy contains all stars in the universe.
 - c. All stars in our Galaxy take the same time to complete one orbit.
 - d. Most stars in our Galaxy are in the central bulge.
 - e. None of the stars in our Galaxy move.
3. What was the Shapley–Curtis debate all about? Was a winner declared at the end of the debate? Whose ideas turned out to be correct?
4. How did Edwin Hubble prove that M31 is not a nebula in our Milky Way Galaxy?
5. What observations led Harlow Shapley to conclude that we are not at the center of the Galaxy?
6. Explain why globular clusters spend more time in the galactic halo than in the plane of the disk, even though their eccentric orbits take them across the disk of the Galaxy.
7. How do hydrogen atoms generate 21-cm radiation? What do astronomers learn about our Galaxy from observations of that radiation?
8. Why do astronomers believe that vast quantities of dark matter surround our Galaxy?
9. Why are there no massive O and B stars in globular clusters?
10. What evidence indicates that a supermassive black hole is located at the center of our Galaxy?
11. What are the least massive galaxies in the universe?
 - a. normal spirals
 - b. barred spirals
 - c. giant ellipticals
 - d. dwarf ellipticals
 - e. irregulars
12. Spiral density waves are directly responsible for which of the following?
 - a. flocculent spirals
 - b. grand-design spirals
 - c. supernovae
 - d. the collisions between galaxies
 - e. galactic cannibalism
13. Which of the following statements about the motion of galaxies is correct?
 - a. All galaxies are moving apart.
 - b. Superclusters of galaxies are all moving apart.
 - c. Superclusters of galaxies are all moving toward each other.
 - d. The Milky Way Galaxy is at the center of the universe.
 - e. All clusters of galaxies in each supercluster are moving toward each other.
14. What is the common name for the sonic boom created by lightning?
15. In spiral galaxies, what spectral classes of stars are only found in spiral arms?

16. What is the Hubble classification scheme? Which category includes the biggest galaxies? Into which category do the smallest galaxies fall? Which type of galaxy is the most common?
17. In which Hubble types of galaxies are new stars most commonly forming? Describe the observational evidence that supports your answer.
18. What is the difference between a flocculent spiral galaxy and a grand-design spiral galaxy?
19. Briefly describe how the theory of self-propagating star formation accounts for the existence of spiral arms in some spiral galaxies.
20. Briefly describe how the spiral density wave theory accounts for the existence of spiral arms in some spiral galaxies.
21. How is it possible that galaxies in our Local Group still remain to be discovered? In what part of the sky are these galaxies located? What sorts of observations might reveal these galaxies?
22. Can any galaxies besides our own be seen with the naked eye? If so, name one.
23. What is the difference between a rich cluster and a poor cluster of galaxies? What is the difference between a regular cluster and an irregular cluster of galaxies?
24. Why do astronomers believe that considerable quantities of dark matter must exist in clusters of galaxies?
25. What is the Hubble law?
26. Some galaxies in the Local Group exhibit blueshifted spectral lines. Why are these blueshifts not violations of the Hubble law?
27. What is a standard candle? Why are standard candles important to astronomers trying to measure the Hubble constant?
28. What kinds of stars would you expect to find populating the space between galaxies in a cluster?
29. What is a double-radio source seen along one axis of a jet called?
 - a. a BL Lac object
 - b. a quasar
 - c. a binary star system
 - d. a double-radio source
 - e. a pulsar
30. What two things does the engine of a quasar contain?
 - a. a supermassive black hole and a binary companion
 - b. a stellar-mass black hole and a binary companion
 - c. a supermassive black hole and an accretion disk
 - d. a stellar-mass black hole and an accretion disk
 - e. a supergiant star and an accretion disk
31. Suppose you suspected a certain object in the sky to be a quasar. What sort of observations would you perform to confirm your hypothesis?
32. What is an active galaxy? List the different kinds of active galaxies. How do they differ from one another?
33. Why do astronomers believe that the energy-producing region of a quasar is very small?
34. How is synchrotron radiation produced?
35. What evidence indicates that quasars are extremely distant active galaxies?
36. What is a double-radio source?
37. What is a supermassive black hole? What observational evidence suggests that supermassive black holes are located at the centers of many galaxies?
38. Why do many astronomers believe that the engine at the center of a quasar is a supermassive black hole surrounded by an accretion disk?
39. How does the orientation of the jets that emanate from the center of a galaxy (relative to our line of sight) relate to the type of active galaxy that we observe?

Advanced Questions

40. Why do astronomers not detect 21-cm radiation from the hydrogen in giant molecular clouds?
41. Describe the rotation curve you would get if the Galaxy rotated like a rigid body.
42. Compare the apparent distribution of open clusters, which contain young stars, with the distribution of globular clusters in the Milky Way. Why are open clusters also referred to as *galactic clusters*?
43. In the 1960s, some astronomers suggested that quasars might be compact objects ejected at high speeds from the centers of nearby ordinary galaxies. Why do astronomers now believe that is not correct? (*Hint: Consider the directions in which those objects would be ejected.*)
44. When quasars were first discovered, many astronomers were optimistic that these extremely luminous objects could be used to probe distant regions of the universe. For example, it was hoped

that quasars would provide high-redshift data from which the Hubble constant could be more accurately determined. Why have these hopes not been realized?

Discussion Questions

45. Discuss the advantages and disadvantages (for example, reliability, repeatability, etc.) of using the various standard candles to determine extragalactic distances.

46. Discuss whether the various Hubble types of galaxies actually represent some sort of evolutionary sequence.

47. Discuss the sorts of phenomena that can occur when galaxies collide. Do you think that such collisions can change the Hubble type of a galaxy? Explain your answers.

48. From what you know about stellar evolution, the interstellar medium, and spiral density wave theory, explain the appearance and structure of the spiral arms of grand-design spiral galaxies.

Got It?

49. Does the Sun move in the Milky Way? If so, describe its motion.

50. About what fraction (or percent) of the matter in the Milky Way can astronomers see with current technology?

51. Where in the Galaxy is the solar system located?

- a. in the nucleus
- b. in the halo
- c. in a spiral arm
- d. between two spiral arms
- e. in the central bulge

52. Is the Milky Way in close contact with other galaxies? Explain.

53. What shape(s) do galaxies have?

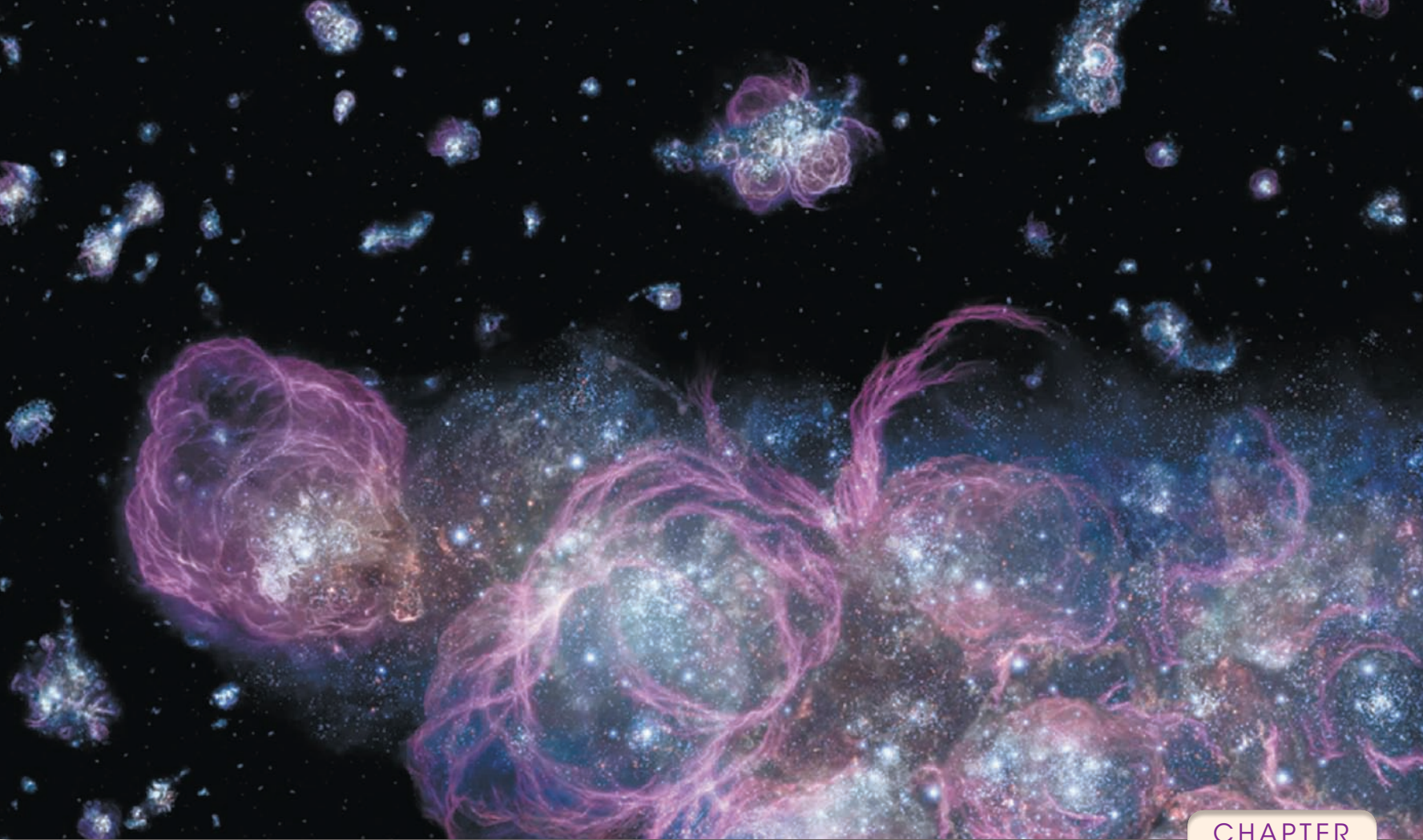
54. With most galaxies in the universe moving away from us equally in all directions, are we at the center of the universe? Explain.

55. Are galaxies uniformly distributed throughout the universe? Explain your answer.

56. Is the Milky Way Galaxy moving through the universe? Justify your answer.

57. What is a quasar?

58. What does a quasar look like through an optical telescope?



CHAPTER 14

Bursts of star formation in the early universe, depicted by science artist Adolf Schaller. (Adolf Schaller, STScI/NASA/K. Lanzetta, SUNY)

Cosmology

WHAT DO YOU THINK?

- 1 What is the universe?
- 2 Did the universe have a beginning?
- 3 Into what is the universe expanding?
- 4 How strong is gravity compared to the other forces in nature?
- 5 Will the universe last forever?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

For millennia, our ancestors speculated about whether the universe has existed forever or whether it began some finite time in the past. Did God create the universe, or was its creation an event that required no divine intervention? Science has yet to provide an answer to this question. But astronomers and other scientists *have* determined that it began at a finite time in the past and they have accumulated a great deal of knowledge about what has happened to the universe starting a tiny fraction of a second after it began and what its fate will be. Although the overall theory of cosmic evolution is not complete, its major elements are consistent with observations.

In this chapter we will examine the distribution of matter on the largest scales in the universe and how this matter is moving and changing. Using this information, we will explore the Big Bang model of how the universe has evolved from its earliest moments

when it was smaller than an atom. We will see how it grew and created matter as we know it, forming and evolving large-scale systems like galaxies and groupings of galaxies. Finally, we will consider the fate of the universe.

In this chapter you will discover

- cosmology, which seeks to explain scientifically how the universe began, how it evolves, and its fate
- the best cosmological theory that we have for the evolution of the universe—the Big Bang
- how astronomers trace the emergence of matter and the formation of galaxies
- how astronomers explain the overall structure of the universe
- our understanding of the fate of the universe

COSMOLOGY

1 The **universe** consists of all matter, energy, and space-time that we can ever detect or that will ever be able to affect us. (We use this definition because there may be matter and energy in other dimensions or matter and energy that are moving away from us so quickly that their influence will never reach us.) So far in this text, we have explored matter on size scales from atoms to superclusters of galaxies. We also learned in Section 13-17 that the superclusters are all moving away from one another, implying that the universe is expanding. In this chapter we take the observational evidence from the rest of the book and use it to explore the Big Bang theory of cosmology. **Cosmology** is the study of the large-scale structure and evolution of the universe.

14-1 General relativity predicts an expanding (or contracting) universe

Modern cosmology almost began in 1915 when Einstein published his theory of general relativity. To his surprise and dismay, the relativity equations predicted that the universe is not static: They indicated that it should be either expanding (which it is) or contracting. But Einstein was not ready for what the equations were telling him.

At the time that Einstein published the theory of general relativity, the existence of galaxies and clusters and superclusters of galaxies had not yet been established and the 1925 discovery of the Hubble flow (Section 13-17) was a decade away. The prediction of a changing universe flew in the face of the then widely accepted belief in an infinite, static universe, a concept promoted by Isaac Newton more than two centuries earlier. Newton

believed that each star is fixed in place and held under the influence of a uniform gravitational pull from every part of the cosmos. If the stars were not uniformly distributed, he argued, one region would have more mass than another. The denser region's gravity would then attract other stars, causing them to further clump together. Because he did not observe this clumping, Newton concluded that the stars in the universe must be distributed uniformly over an infinite space.

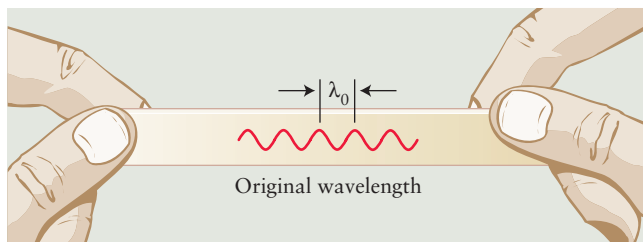
The apparently static universe and the prevailing philosophy that the universe had existed forever made Einstein doubt the implications of his own theory, so he missed the opportunity to propose that we live in a changing universe. Instead, he adjusted his elegant equations to yield a static, finite cosmos. He did this by adding a repulsive (outward-pushing) term, called the **cosmological constant**, to his equations so that gravity's normal attractive force would be counterbalanced and the universe would be static. After observations revealed that the universe is expanding, Einstein said that adding the cosmological constant was the biggest blunder of his career.

Although the value of the cosmological constant that Einstein inserted was wrong, the concept of such a constant may be correct. Observations since 1997 indicate that the universe is not just expanding but actually *accelerating* outward. This acceleration means that there must be an outward pressure that more than counteracts the effects of normal gravitation, which is trying to slow the universe's expansion. One of the two current theories that can explain this acceleration is the presence of a cosmological constant that creates the outward pressure. We discuss these two theories further in Section 14-15.

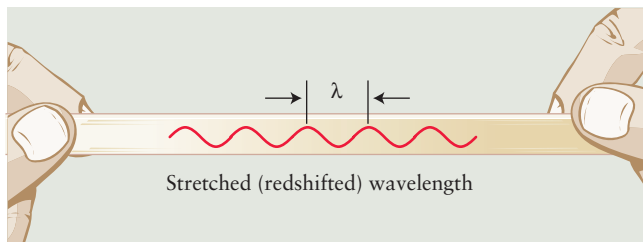
14-2 The expansion of the universe creates a Dopplerlike redshift

Edwin Hubble is credited with discovering that we live in an **expanding universe** (see Section 13-17). The redshifts of clusters and superclusters of galaxies that Hubble found moving away from us appear to be produced by the Doppler effect, but they actually are not. Recall that the normal Doppler shift is caused by an object moving toward or away from us through *fixed* spacetime (see Sections 3-2 and 3-18). However, using Einstein's theory of general relativity, we find that spacetime, the fabric of the universe, is *not fixed* but is actually expanding. This expansion is what is carrying the superclusters away from each other and, in many cases, carrying clusters in a given supercluster away from one another. (The gravitational force that holds objects like planets and stars and entire galaxies together is so strong that these bodies and systems are not being pulled apart. The expansion of space just acts to increase the separation between superclusters of galaxies and between clusters of galaxies in the same supercluster that are not gravitationally bound to each other.)

The redshift that Hubble observed, caused by the expansion of the universe, is properly called the **cosmological redshift**. In other words, the photons that we observe from galaxies in other superclusters are all redshifted because space is expanding. To understand why, consider a wave drawn on a rubber band (Figure 14-1). The wave has an unstretched wavelength (see Section 3-1) of λ_0 . As the rubber band stretches, the wavelength increases ($\lambda > \lambda_0$). Now imagine a photon coming toward



a A wave drawn on a rubber band ...



b ... increases in wavelength as the rubber band is stretched.

FIGURE 14-1 Cosmological Redshift Just as the waves drawn on this rubber band are stretched along with the rubber band, so too are the wavelengths of photons stretched as the universe expands.

us from a distant galaxy. As the photon travels through space, space is expanding, and, like stretching the rubber band, this expansion stretches the photon's wavelength. When the photon reaches our eyes, we see a drawn-out wavelength—the photon is redshifted. The longer the photon's journey, the more its wavelength is stretched by the expansion of the universe. Therefore, astronomers observe larger redshifts in photons from relatively distant galaxies than in photons from relatively nearby galaxies.

As we saw in Section 13-16, Hubble discovered the linear relationship between the distances to galaxies in other superclusters and the redshifts of those galaxies' spectral lines. For example, a galaxy twice as far from Earth as another galaxy has twice the cosmological redshift of the closer one. The normal Doppler shift has the same relationship. As a result of the two effects being described by the same equations, the normal Doppler shift and the cosmological redshift predict the same relationship between redshift and motion—except for the most distant galaxies and quasars, where effects of relativity must be taken into account.

Working with relatively nearby galaxies, Hubble was fully justified in using the “normal” Doppler equation to calculate the recession of galaxies.

Focus Question 14-1

What are two kinds of redshifts presented in this book?

14-3 The Hubble constant is related to the age of the universe

Hubble's law gives us a way to estimate the age of the universe. Imagine watching a movie of two superclusters receding from each other. If we then run the film backward, time runs in reverse, and we observe all the superclusters coming together. The time it would take for them all to collide is the time since they were last combined as a single clump of matter. *Assuming that the universe began expanding when it came into existence and that it has always expanded at a constant rate*, we can use a simple equation to estimate the age of the universe:

$$\text{Time since the Big Bang} = \frac{\text{separation distance}}{\text{recessional velocity}}$$

Recall that Hubble found the relationship

$$\text{Recessional velocity} = H_0 \times \text{separation distance}$$

which we can rewrite as

$$H_0 = \frac{\text{recessional velocity}}{\text{separation distance}}$$

Comparing the first and last equations here, we see that Hubble's constant is the reciprocal of the time since the universe began. Using a Hubble constant of 69 km/s/Mpc, we find:

$$1/H_0 \approx 1/69 \text{ km/s/Mpc} \approx 13.8 \text{ billion years}$$

As of 2014, the most accurate measurement is 13.798 ± 0.037 billion years. In what follows, we will simply use an age of 13.8 billion years.

THE BIG BANG

2 3 In 1927, the Belgian astrophysicist and Catholic priest Georges Lemaître proposed that, as a consequence of the equations of Einstein's theory of general relativity, the universe is expanding. Using Hubble's subsequent observations of an expanding universe, Lemaître logically concluded that the superclusters must have expanded from a much smaller volume. He proposed that the universe began as an extraordinarily hot and dense *primordial atom* of energy. Just as energy created in an explosion causes debris to expand outward, the primordial energy created at the beginning of time caused the universe to expand. This initial event is now called the **Big Bang**. Space and time (that is, spacetime; see Section 12-11) did not exist before that moment. Rather than expanding into preexisting spacetime, the Big Bang explosion

created spacetime, and that spacetime (in which we live) has been expanding ever since. The Big Bang led to the formation of all spacetime, matter, and energy. As noted earlier in this chapter, science does not yet have an explanation for why the Big Bang occurred.

Shortly after World War II, the astrophysicists Ralph Alpher and George Gamow proposed that a fraction of a second after the Big Bang, the universe was an incredibly hot blackbody—in excess of 10^{32} K. As heat is a measure of how many photons there are and how energetic they are, the early universe must therefore have been filled with intensely high-energy electromagnetic radiation. As the universe expanded, the cosmological redshift stretched these gamma rays to longer and longer wavelengths so that most of the photons left over from this event are now radio waves. Recall from Section 3-3 that photon energies decrease with increasing wavelength. Because the early photons are being redshifted, the energy they each have is decreasing. This energy is measured as heat, so the universe is cooling off as it expands.

14-4 Remnants of the Big Bang have been detected

Calculations indicated that if the universe began with a hot Big Bang, the remnants of that energy should still fill all space today, giving the entire universe a temperature only a few kelvins above absolute zero. This radiation is called the **cosmic microwave background** because the peak of its blackbody spectrum should lie in the microwave part of the radio spectrum. In the early 1960s, Robert Dicke, P. J. E. “Jim” Peebles, David Wilkinson, and

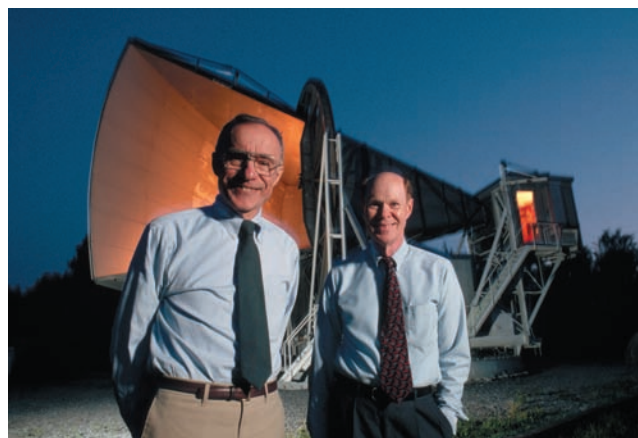


FIGURE 14-2 Bell Labs Horn Antenna This Bell Laboratories horn antenna at Holmdel, New Jersey, was used by Arno Penzias (right) and Robert Wilson in 1965 to detect the cosmic microwave background. This discovery earned them the Nobel prize in 1978. (Roger Ressmeyer/CORBIS)

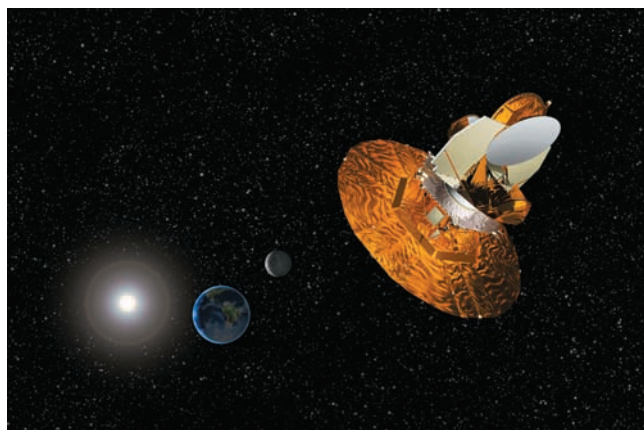
their colleagues at Princeton University began designing a microwave radio telescope to detect it. Meanwhile, just a few miles from Princeton, two physicists had already detected the cosmic background radiation. Arno Penzias and Robert Wilson of Bell Laboratories were working on a new horn antenna designed to relay telephone calls to Earth-orbiting communications satellites (**Figure 14-2**). However, Penzias and Wilson were mystified. No matter where in the sky they pointed their antenna, they detected a faint background noise. All efforts to eliminate this background noise failed, even the careful removal of static noise-generating pigeon droppings from the antenna.

Thanks to a colleague, they learned of the then-theoretical cosmic microwave background and the work of Dicke, Peebles, and Wilkinson in trying to locate it. Communicating with the Princeton astronomers, Penzias and Wilson presented their finding and thus were able to claim the first detection—their annoying noise was, in fact, the remnant energy of the Big Bang. The cosmic microwave background is required by the Big Bang theory but is neither required nor predicted by any other competing cosmological theory. Detection of the cosmic microwave background is the principal reason why the Big Bang is accepted by astronomers as the correct cosmological theory.

Precise measurements of the cosmic microwave background were first made by the *Cosmic Background Explorer* (COBE) satellite, which operated between 1989 and 1994. The radiation has since been measured more precisely by other equipment, such as the *Wilkinson Microwave Anisotropy Probe* (WMAP) and the balloon-carried *BOOMERANG* (**Figure 14-3**). The data taken by COBE and shown in **Figure 14-4** reveal that, as predicted, this ancient radiation has the spectrum of a

Focus Question 14-2

If the universe were twice as old as it is now, how would the Hubble constant compare to the value it has today?



a



b

RIVUXG

FIGURE 14-3 In Search of Primordial Photons (a) The *Wilkinson Microwave Anisotropy Probe* (WMAP) satellite, launched in 2001, improved upon the measurements of the spectrum and angular distribution of the cosmic microwave background taken by the COBE satellite. (b) The balloon-carried telescope BOOMERANG orbited above Antarctica for 10 days, collect-

ing data used to resolve the cosmic microwave background, with 10 times higher resolution than that of COBE. All of these experiments found local temperature variations across the sky but no overall deviation from a blackbody spectrum. (a: NASA/WMAP Science Team; b: BOOMERANG/NASA/NSF)

blackbody with a temperature of approximately 2.73 K (referred to as the *3-degree background radiation*). The observations by COBE earned the 2006 Nobel Prize in Physics for astrophysicists John C. Mather and George F. Smoot.

The photons of the 3-degree background radiation are not the only remnants of the Big Bang that astronomers have found. In 2008, neutrinos created along with the photons in the Big Bang were indirectly observed by WMAP. (Recall from Section 9-7 that neutrinos are also created today in the nuclear fusion occurring in the Sun and other stars.) The discovery of these *cosmic*

neutrinos was made by comparing the predicted effects of those neutrinos on the young universe with observations of the earliest observable remnants of the Big Bang, about which we will have more to say shortly. The theory of these neutrinos and the observations of their predicted effects agree to high precision. The numbers of neutrinos in the *cosmic neutrino background* is impressive: Several million of them pass through your body every second. There are similar numbers of neutrinos in similar volumes of space everywhere in the universe.

Focus Question 14-3

What other objects that we have studied in this book have (nearly) blackbody spectra?

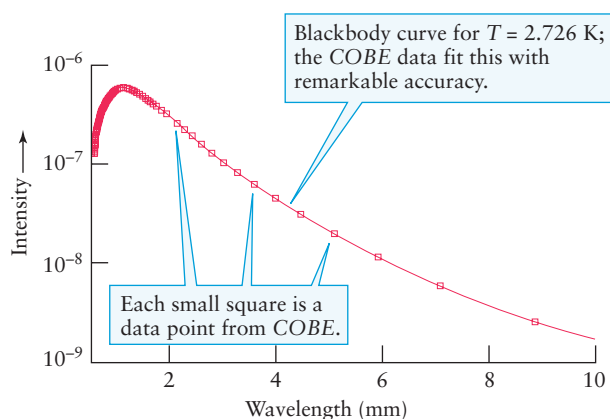


FIGURE 14-4 Spectrum of the Cosmic Microwave Background The little squares on this graph are COBE's measurements of the brightness of the cosmic microwave background plotted against wavelength. To a remarkably high degree of accuracy, the data fall along a blackbody curve for 2.73 K. The peak of the curve is at a wavelength of 1.1 mm, in accordance with Wien's law. (E. Cheng; NASA COBE Science Team)

Insight Into Science

Consider Plausible Alternatives Until They Are Shown to Be Inconsistent with Observations Sir Fred Hoyle, who aptly named the Big Bang in 1949, was one of the very few astrophysicists who never believed that it occurred. Hoyle, Herman Bondi, and Thomas Gold proposed the *steady-state theory* in which the universe has existed forever, is continuously expanding, and, as it expands, new matter is created that takes the place of the receding matter. Where this new matter might come from is unknown, but the laws of physics do not necessarily exclude its creation. Today, astronomers accept the Big Bang theory rather than the steady-state theory because a cosmic microwave background is predicted to exist by the Big Bang theory but not by the steady-state theory.

14-5 The universe has two symmetries— isotropy and homogeneity

Observations show that the cosmic microwave background is almost perfectly isotropic: Its intensity is nearly the same in every direction in the sky. Indeed, if you could look at the cosmic microwave background from all directions in the entire universe with the naked eye, it would appear as uniform as a well-painted wall. However, when extremely high-resolution measurements are made, tiny differences in the temperature of the universe are observed in different directions. One difference is caused by our motion through the cosmic microwave background.

Figure 14-5 shows a map of the microwave sky. It is very, very slightly warmer than average in the direction of the constellation Leo and slightly cooler in the opposite direction, toward Aquarius. The observed temperature differences shown in Figure 14-5 mean that the solar system is moving toward Leo at a speed of 390 km/s. These observations can be understood as follows: As we move toward Leo, our motion through the cosmic microwave background radiation creates a normal Doppler shift that causes photons from that direction to appear to have shorter wavelengths—radiation from the direction of Leo is blueshifted, whereas radiation from the opposite direction, Aquarius, is redshifted because we are moving away from that region (**Figure 14-6**). Furthermore, a photon's wavelength determines its energy (see Section 3-3). The energy of the photons then determines the temperature that we measure for that part of the sky. Observing slightly shorter-wavelength photons in the direction of Leo means that space in that direction appears slightly warmer, because of our motion, than does space in other directions.

As noted in Section 13-14, the Milky Way and the rest of our Local Cluster are part of the Laniakea supercluster. Taking into account the known velocity of the Sun around the center of our Galaxy, astronomers calculate that the entire Milky Way Galaxy is moving relative to the cosmic microwave background at 600 km/s—some 1.3 million mi/h—in the general direction of the Virgo cluster. The gravitational attraction of four nearby clusters of galaxies, most significantly the *Great*

Attractor, and an enormous supercluster called the *Shapley concentration*, located behind the Virgo cluster from our perspective, is believed to be pulling us in that direction. Subtracting the motions of Earth, the Sun, the Milky Way, and the Local Group (of galaxies)

with respect to the microwave background, the average intensity of radiation is found to be the same in all directions (that is, isotropic) to about 1 part in 100,000.

Although the existence of the cosmic background radiation provides compelling support for the Big Bang theory, its near-isotropy was, until recently, a major problem. The issue was that the original Big Bang theory did not require space to be isotropic—after accounting for Earth's mo-

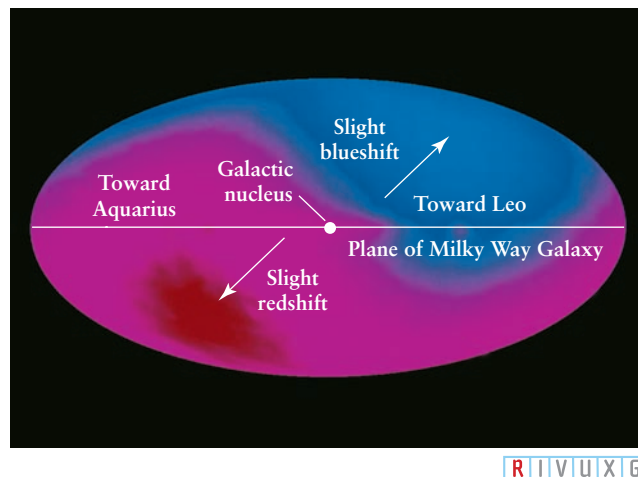


FIGURE 14-5 Microwave Sky This map of the microwave sky was produced from data taken by instruments on board COBE. The galactic center is in the middle of the map, and the plane of the Milky Way runs horizontally across the map. Color indicates Doppler shift and temperature: Blue is where the microwave background is blueshifted and appears warmer, while red is where it is redshifted and appears cooler. This Doppler shift across the sky is caused by Earth's motion through the microwave background. The resulting variation in background temperature is quite small, only 0.0033 K above the average radiation temperature of 2.726 K (which we have rounded to 2.73 K in the text). (NASA)

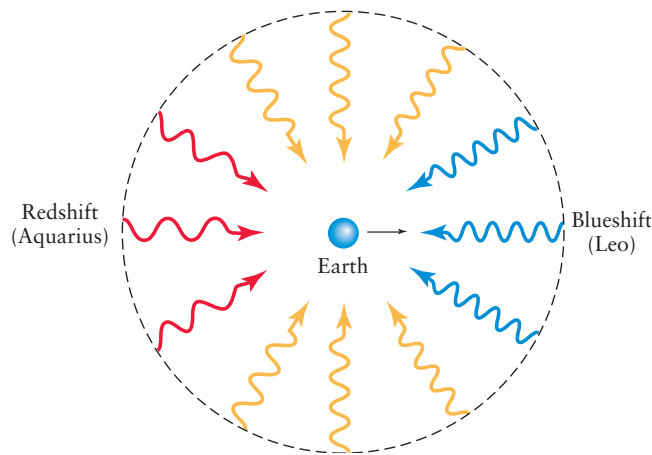


FIGURE 14-6 Our Motion Through the Microwave Background Because of the Doppler effect, we detect shorter wavelengths in the microwave background and a higher temperature of radiation in that part of the sky toward which we are moving. This part of the sky is the area shown in blue in Figure 14-5. In the opposite part of the sky, shown in red in Figure 14-5, the microwave radiation has longer wavelengths and a cooler temperature.

tion through the cosmos, the temperature in one direction *could* have been many degrees different than the temperature in any other direction, but it is not.

Isotropy is not limited to the blackbody radiation observed throughout the universe. It is also found on sufficiently large scales when exploring the numbers of

Focus Question 14-4

What object other than Earth could meaningfully have been used at the center of Figure 14-6?

galaxies in different directions. Astronomers have counted the numbers of galaxies at different distances from us and in a variety of directions. The number of galaxies stays roughly constant with respect to both distance and direction, allowing, as we saw in Section 13-14, for the distribution of most galaxies on the boundaries of sponge-shaped regions or bubbles throughout the universe. Because of this *uniformity with distance*, we say that the universe is also homogeneous. Isotropy and **homogeneity** must be explained by any viable theory of cosmology, such as the Big Bang.

The need to explain isotropy and homogeneity, among other things, has led to numerous refinements in the Big Bang theory. As a result, this theory now provides an accurate scenario for the evolution of the universe from a tiny fraction of a second after it formed and onward.

A BRIEF HISTORY OF SPACETIME, MATTER, ENERGY, AND EVERYTHING

The composition of the universe during its first few minutes of existence was profoundly different from what it is today. To understand why, we need to expand briefly on the nature of the four known physical forces: gravity, electromagnetism, and the strong and weak nuclear forces.

14-6 All physical forces in nature were initially unified

Only the effects of electromagnetism and gravity can extend over infinite distances. The electromagnetic force holds electrons in orbit around the nuclei in atoms, which allows normal matter to exist. However, over large volumes of space, the net effects of electromagnetism cancel out because a negative electric charge exists for every positive charge and a south magnetic pole exists for every north magnetic pole.

4 Although gravity is the weakest of all four forces, it is the only force whose effect is both infinite in extent and attractive for all matter and energy. Those characteristics are why gravity determines the evolutionary behavior of stars, the orbits of planets, the existence of galaxies, and many other large-scale phenomena. Indeed, the attractive force of gravity dominates the universe at astronomical distances. (Despite all of this, we will learn shortly that, under certain conditions, gravity may have a repulsive component.)

In contrast to gravity and electromagnetism, both the strong and the weak nuclear forces act over extremely short ranges. Their influences extend only over atomic nuclei, distances less than about 10^{-15} m. The **strong nuclear force** holds protons and neutrons together. Without this force, all nuclei other than hydrogen would burst apart because of the electromagnetic repulsion of their positive-

ly charged protons. Thus, the strong nuclear force overpowers the electromagnetic force inside nuclei.

The **weak nuclear force** is at work in certain kinds of radioactive decay, such as the transformation of a neutron into a proton. Protons and neutrons are composed of more basic particles, called **quarks**. A proton is composed of two “up” quarks and one “down” quark, whereas a neutron is made up of two “down” quarks and one “up” quark. The weak nuclear force is at play whenever a quark changes from one variety to another. For example, in one kind of radioactive decay, a neutron transforms into a proton as a result of one of the neutron’s “down” quarks changing into an “up” quark.

To examine details of the physical forces, scientists use particle accelerators that hurl high-speed particles at targets. High speed is equivalent to high energy and therefore high temperature. In such experiments, physicists find that *at sufficiently high temperatures, the different forces begin to behave the same way*. For example, at extremely high temperatures, the electromagnetic force, which works over all distances under “normal” circumstances, and the weak force, which only works over very short distances under the same “normal” circumstances, become identical. They are no longer separate forces, but become a single force called the *electroweak force*. The experiments verifying this were done at the CERN particle accelerator in Europe in the 1980s.

Theories have also been developed that describe conditions under which the electromagnetic, weak, and strong forces all have the same strength (that is, when they act as one). These **Grand Unified Theories** (or GUTs) describe how the energies or, equivalently, temperatures at which this unification of forces occurs are greater than can ever be achieved in any laboratory on Earth. But recall from the earlier discussion that when the universe was young, it was very hot—hot enough for this unifying of forces to occur. Detailed calculations reveal that in the first 10^{-35} seconds of existence, the temperature of the universe was more than high enough for the electromagnetic, weak, and strong forces to have the same strength.

The relationship between gravitation and the other three forces is still uncertain. Scientists have not yet been able to include gravity’s effects in a single theory with the weak, strong, and electromagnetic forces. Efforts to create such a theory are under way, under the general name **Theory of Everything**. Mathematical formulations of this theory, including loop quantum gravity, **superstring theory**, and M-theory, are being developed, although these theories do not yet make predictions that scientists have been able to test, putting them on the outskirts of science (see Discovery 14-1, at the end of this chapter, which provides an introduction to superstring theory). Assuming gravity will be joined to the other forces, we include its expected effects here. **Figure 14-7** shows that, during the beginning moments of the universe (from 0 to 10^{-43} seconds, called the **Planck era**), the four forces are believed

Focus Question 14-5

Is the space in a sphere 3 m (3 yds) around you isotropic?

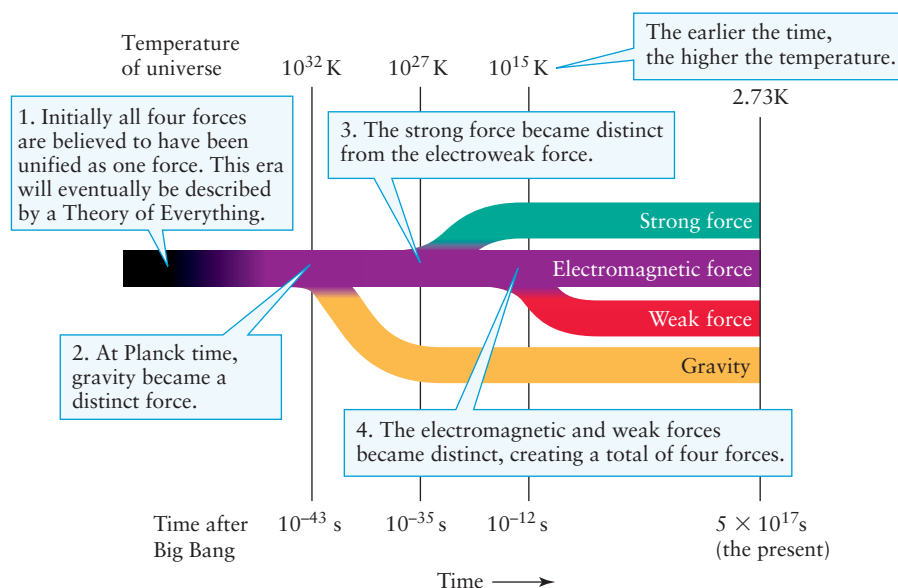


FIGURE 14-7 Unification of the Four Forces The four physical forces were initially a single force. This symmetry was broken as the universe expanded and cooled. This figure shows the time and temperature of the universe when the forces separated from each other.

to have all been unified—only one physical force existed in nature. However, as it assumes a Theory of Everything force exists, this assumption remains to be proven.

Insight Into Science

Different Realms of Science Working Together

Because physicists who study elementary particles have no hope of building accelerators powerful enough to test the unified theories of fundamental forces described here, they have joined with astrophysicists and astronomers to use the entire universe as a large laboratory. If there is a way to see, even indirectly through remnants of events, what happened during the earliest moments of the universe, they could use those observations to test the theory. Amazingly, such observations can be, and are being, made.

14-7 Equations explain the evolution of the universe, even before matter and energy, as we know them, existed

The earliest time at which our current equations can explain the behavior of the universe is about 10^{-43} s after the Big Bang. This is called the **Planck time**, in honor of Max Planck, who helped derive some of the fundamental quantum physics that was used to calculate this time. The Planck time denotes the end of the Planck era. At the Planck time, gravity is thought to have become a separate force, leaving electromagnetism and the weak and strong nuclear forces still united as the GUT force. Which of the present GUT models, if any, is correct, remains to be seen.

The Planck time corresponded to when the currently observable universe was much smaller than the size of an atom today. Keep in mind that the universe was not (and is not) expanding into preexisting spacetime. Rather, the

expansion was (and is still) creating spacetime. Therefore, you could not stick a thermometer into the early universe from “outside” to take its temperature. However, we can calculate the total amount of energy in the universe at the Planck time, and it corresponds to a temperature then of about 10^{32} K. There were no separate particles and electromagnetic radiation then, just gravitation and the GUT force, so matter as we now know it did not exist.

During the first 10^{-35} seconds of the universe’s existence (the blink of an eye takes about one-tenth of a second), calculations indicate that the expanding universe cooled to 10^{27} K (Figure 14-8). (For comparison,

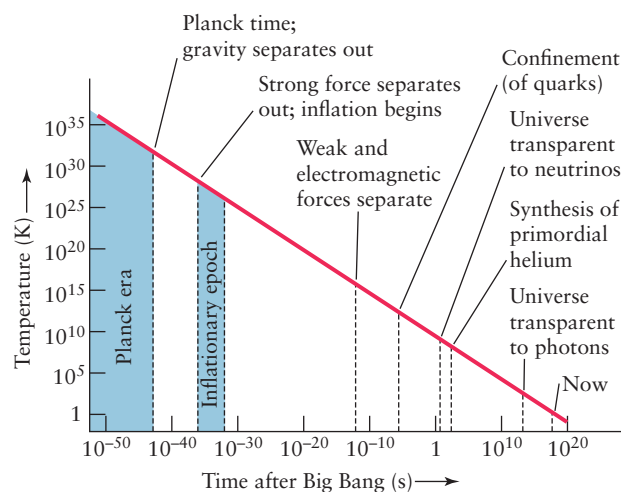


FIGURE 14-8 Early History of the Universe Current theory holds that, as the universe cooled, the four forces separated from their initially unified state. The inflationary epoch lasted from 10^{-36} seconds to 10^{-32} seconds after the Big Bang. Quarks became confined together, thereby creating neutrons and protons 10^{-6} seconds after the Big Bang. The universe became transparent to light (that is, photons decoupled from matter) when the universe was about 1.6×10^{13} seconds (380,000 years) old. The physics of the Planck era is presently unknown.

room temperature is about 300 K.) By then, the energy in each volume of the universe had decreased to the point at which the strong nuclear force could no longer remain unified with the electromagnetic and weak nuclear forces. At that instant, the strong nuclear force emerged (see Figure 14-7), distinct from the other two forces, which remained together as the electroweak force. For a short period, then, there were three forces in nature: gravitation, the strong nuclear force, and the electroweak force.

Many astronomers hypothesize that during the time that the strong force was decoupling from the electroweak force, the universe was in an unstable state, called a *false vacuum*. It was unstable in the same sense as you would be if you woke up one day to find yourself on a greased, half-meter (twenty-inch) wide shelf on the face of a cliff thousands of meters above the cliff bottom. The slightest move and you slide off and fall. Analogous to the transition you make by falling off the shelf, the young universe is believed to have suddenly made a transition (Figure 14-9) into a stable *true vacuum* state. The universe's transition to the true vacuum state caused its expansion to momentarily accelerate.

As the universe began moving toward the true vacuum state, it was already expanding. The transition caused the expansion rate to increase dramatically. From about 10^{-36} seconds until about 10^{-32} seconds after the Big Bang, the universe ballooned outward about 10^{50} times larger (Figure 14-10). The currently observable universe expanded to about the size of a soccer ball. This huge rate of expansion of the universe is termed **inflation**. Shortly after inflation ended, at about 10^{-12} seconds after the Big Bang, the weak and electromagnetic forces became separate, heralding the first formation of elementary particles: quarks, electrons, photons, neutrinos, and their antiparticles.

14-8 Homogeneity and isotropy are results of inflation

Because disorder is more likely to occur than order in nature, astronomers theorize that the matter and energy in the early universe were distributed chaotically and expanded turbulently. That is also why an explosion on Earth expands turbulently, with lots of swirling gases, rather than expanding smoothly, like the surface of a balloon as it is being inflated. As a result, some places in the early universe had higher temperatures than others. If the universe had never undergone inflation, but had continued expanding only at the rate calculated for its motion at the Planck time, today we should be able to see different parts of the universe with significantly different temperatures. However, as we saw in Figure 14-5, the temperatures of every region of the universe are

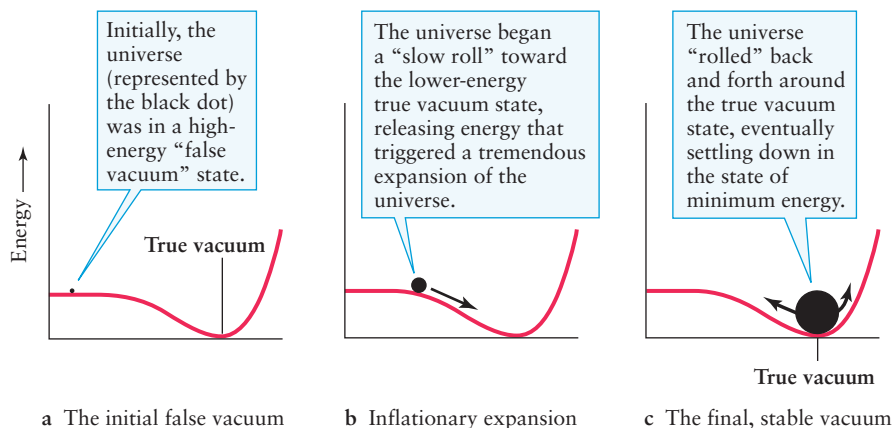


FIGURE 14-9 Cause of Inflation (a) The universe formed in an unstable energy state that (b) began to transition to a stable configuration. (c) This transition provided the energy that caused inflation.

the same to within a very tiny fraction of 1 K. This situation is called the **isotropy problem** or **horizon problem**. Realizing how uniform the universe's temperature is, astronomers have hypothesized that inflation very early in its existence smoothed out the universe that we can see.

During the so-called **inflationary epoch**, the universe became so large that today we can see only a tiny portion of it. We are limited in what we can see by the speed of light. Recall that light takes time to travel

Focus Question 14-6

Which fundamental force is essential in the process of nuclear fusion?

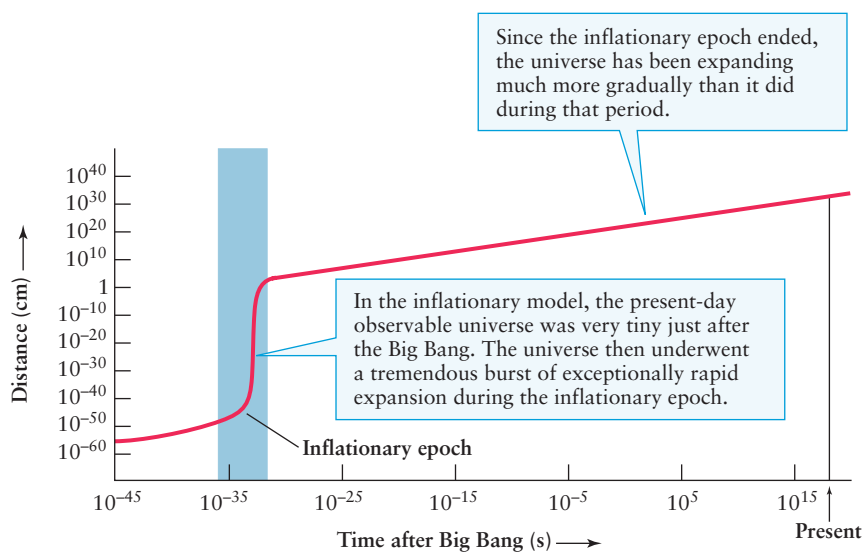


FIGURE 14-10 Observable Universe Before and After Inflation Shortly after the Big Bang, the universe expanded by a factor of about 10^{50} due to inflation. This growth in the size of the presently observable universe occurred in a very brief time (shaded interval).

from astronomical objects to Earth. We see the Sun, for example, as it was about 8 minutes ago. The farther we look into space, the further back in time we are seeing. Because the universe came into existence about 13.8 billion years ago, we cannot see objects that are farther than 13.8 billion light-years away from Earth (see the caveat about this number in the caption to Figure 14-11).

At a distance of 13.8 billion light-years, we see objects like galaxies and clusters of galaxies as they were just being formed. But because of inflation, the matter comprising these apparently distant objects was once extremely close to us. At that early time, before inflation, these objects had the same temperature, pressure, and density as the matter and energy in our nearest neighborhood of the cosmos. As the universe inflated, that nearby matter and energy became much more distant but maintained

Focus Question 14-7

Why is inflation a necessary part of the Big Bang theory of cosmology?

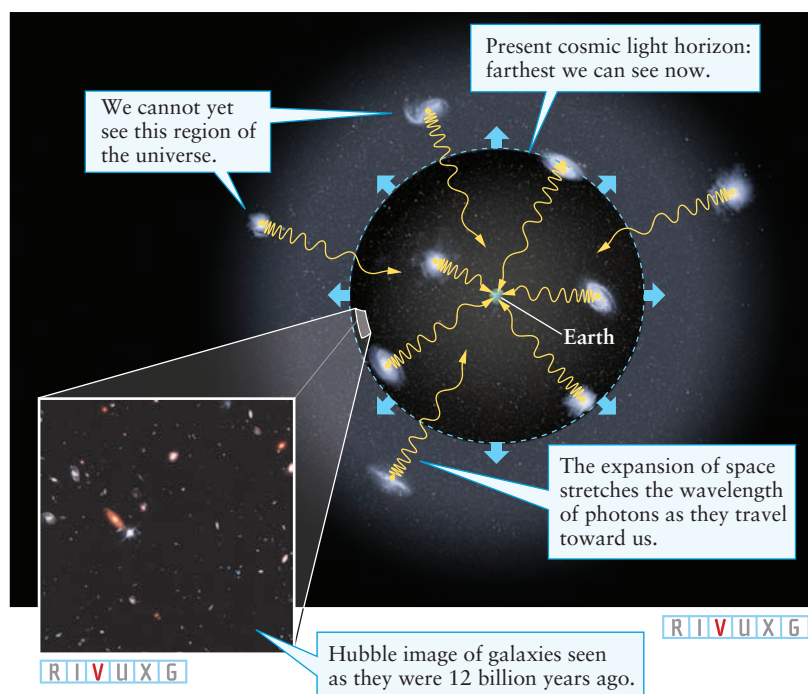


FIGURE 14-11 Observable Universe This diagram shows why we see only part of the entire universe. As time passes, this volume grows, meaning that light from more distant galaxies reaches us. The farthest galaxies we see appear as they were within a few hundred million years after the Big Bang. These galaxies, formed at the same time as the Milky Way, appear young because the light from their beginnings is just now reaching us. While the light from the most distant galaxies we see was traveling toward us, the universe has been growing. Therefore, objects that appear 13 billion light-years away from us are actually about 3 times farther away today. Put another way, if we could see it all as it is today, rather than as it was when photons from distant objects started their journeys, the visible universe would be nearly 46 billion light-years in radius. **Inset:** This image shows some of the most distant galaxies we have seen. (inset: Robert Williams and the Hubble Deep Field Team [STScI] and NASA)

those properties. Therefore, the most distant reaches of space we can see are composed of matter and energy with the same characteristics as in our present neighborhood of the universe. That is why the visible universe appears uniform (homogeneous and isotropic).

Despite being able to see so far away, there is still much of the universe we cannot see. All that we can see from Earth is contained in an enormous sphere (Figure 14-11). The boundary of this sphere is called the **cosmic light horizon**. Light from all of the luminous objects in this volume has had time to reach Earth, but light from objects beyond the cosmic light horizon has not yet arrived here, so we are unaware of objects in these regions. As time goes on, we will be able to see objects farther from us.

14-9 During the first second, most of the matter and antimatter in the universe annihilated each other

When inflation ended, the universe resumed the relatively leisurely expansion initiated by the Big Bang (see Figure 14-10) that continues today as the Hubble flow (see Section 13-17). Arriving at the time 10^{-12} seconds, when the temperature of the universe had dropped to 10^{15} K, the electromagnetic force separated from the weak nuclear force (as noted in Section 14-7; see also Figures 14-7 and 14-8). It is also likely that dark matter formed by this time, but because we know virtually nothing of its nature, we can only acknowledge its existence. From that moment on, all four forces interacted with particles essentially as they do today. At 10^{-12} seconds, the universe's matter was primarily the individual building blocks of protons and neutrons—namely, quarks, antiquarks, and the other elementary particles: neutrinos, antineutrinos, electrons, and positrons.

The next significant event is calculated to have occurred at 10^{-6} seconds, when the universe's temperature dropped to 10^{13} K. Just as it is energetically favorable for boiling water to turn into gas, the temperature of the universe was so high before this time that it was energetically favorable for quarks to exist as isolated particles, rather than to combine in twos and threes to create the particles that exist in the universe today. We know they were initially isolated because similar seas of quarks were recreated under equivalent conditions in the laboratory in 2001 by smashing dense atoms together. After 10^{-6} seconds, a period appropriately called **confinement**, the universe was sufficiently cool so that quarks could finally stick together to form individual protons, neutrons, and their antiparticles. Figure 14-8 summarizes the connections between particle physics and cosmology.

The photons in the early universe contributed to the existence of particles back then. Equations show

that early in the first second, the universe was so hot that virtually all photons were gamma rays possessing incredibly high energies. These energies were high enough so that photons could be transformed into matter and antimatter, according to Einstein's equation $E = mc^2$. This process is the reverse of what we saw in Section 9-7, where we discussed how matter is converted into energy in nuclear fusion. To make a particle of mass m , you need a photon with energy E at least as great as mc^2 , where c is the speed of light.

The creation of matter from energy is routinely observed in laboratory experiments that involve high-energy gamma rays. A highly energetic photon colliding with an atomic nucleus can create a pair of particles (Figure 14-12a). This process, called **pair production**, always creates one ordinary particle and its antiparticle. For example, if one of the particles is an electron, the other is an antielectron, also called a positron. (Recall from Section 12-18 that similar pairs of particles are also believed to be produced near a black hole.)

A positron has all the same properties as an electron except for having an opposite charge. When an electron and a positron meet, they annihilate each other and their energy is converted into photons (Figure 14-12b). Theory predicts that at times earlier than 1 second in the evolution of the universe, photon collisions created vast numbers of pairs of particles. Consequently, shortly after the Big Bang, all space was chock-full of protons, neutrons, electrons, neutrinos, their antiparticles, and dark matter, all immersed in a phenomenally hot bath of high-energy photons.

As the universe expanded, the temperature decreased; after the first second it was sufficiently cool so that photons could no longer create pairs of particles and antiparticles. Particles and antiparticles continued to collide, however, annihilating each other and converting their masses back into high-energy photons, but

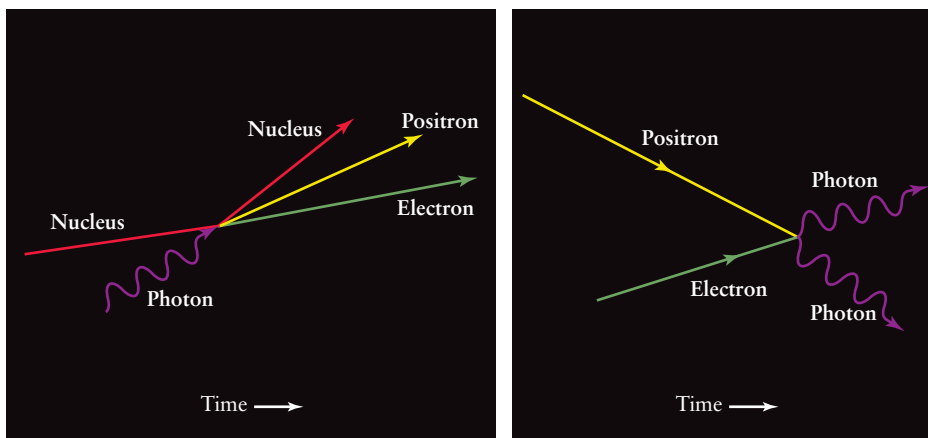
nothing could replenish the dwindling supply of particles and antiparticles.

If all particles had annihilated their antiparticles in the early universe, no matter would be left at all and we would not be here. And yet, here we are and astronomers observe very few antiparticles in the universe (see, for example, Section 13-4). Because particles and antiparticles are always created or destroyed in pairs, why aren't they found in equal numbers in the universe today, or rather, why didn't they all destroy each other shortly after they were created?

Physicists theorize that the symmetry or equality between the number of particles and the number of antiparticles in the early universe was broken by the weak force—a *symmetry breaking*. This effect has been observed on Earth in high-energy particle accelerators over the past half century. Due to symmetry breaking, the number of particles was actually slightly greater than the number of antiparticles in the early universe. For every billion antiprotons, perhaps a billion plus one protons formed; for every billion antielectrons, a billion plus one electrons formed. Virtually all of the antiparticles created in the first second of time annihilated normal particles shortly thereafter. The remaining particles in the universe today are the slight excess of normal matter created as a result of this symmetry breaking.

14-10 The universe changed from being controlled by radiation to being controlled by matter

As the universe expanded and cooled further, the remaining protons and neutrons began colliding and fusing together, a process called **primordial nucleosynthesis**. For the most part, the nuclei with two or more particles were quickly separated again by the gamma rays in which they were bathed, a process called nuclear fission. However, during



a Nucleus collides with photon and creates positron-electron pair

b Positron and electron annihilate each other and create two photons

FIGURE 14-12 Pair Production and Annihilation (a) A particle and an antiparticle can be created when a high-energy photon collides with a nucleus. (b) Conversely, when a particle and an antiparticle annihilate each other, their masses are converted into energy in the form of gamma rays. (The processes are more complex and are only summarized in these drawings.)

Insight Into Science

Share Information Although the evolution of the universe is a truly large-scale event, the activity in the first 3 minutes was nevertheless intimately connected to the creation and properties of elementary particles, the smallest objects in the universe. Physicists who study elementary particles and astrophysicists who study cosmology now regularly hold meetings to compare notes and connect their fields.

the first 3 minutes, enough helium and lithium atoms were created (and not subsequently split apart) to form most of the free helium and at least 10% of the trace element lithium that exist today. A lithium atom has three protons and three neutrons. Hydrogen, helium, and lithium are the three lowest-mass elements. After a few minutes, the universe was too cool to allow fusion to create more massive elements, such as carbon and oxygen. All elements other than the three lowest-mass ones formed later, due to stellar evolution, as discussed in Chapters 11 and 12.

Calculations indicate that for tens of thousands of years, photons dominated the behavior of the universe in two important ways. One effect of the photons was to contribute to the gravitational attraction in the universe. This contradicts intuition, because photons are massless and mass is what we normally associate with causing gravitational attraction. Nevertheless, the gravitational contribution of photons to the universe is a tested prediction of general relativity. For about 30,000 years after the Big Bang, the energy density, and hence the gravitational effect, of photons was greater than the gravitational effect of matter. This situation was called the **radiation-dominated universe**.

The major effect of the radiation-dominated universe was that photons prevented matter from collapsing together and forming stars and galaxies. As the universe expanded, the gravitational effects of both photons and matter decreased as they became more spread out. The effect of photons decreased faster than that of matter (Figure 14-13) because the photons lost energy, and hence gravitational influence, due to the cosmological redshift, but particles with mass did not similarly lose mass or energy. After about 30,000 years, matter came to dominate the gravitational behavior of the universe, a situation that has persisted ever since. As a result, we live in a **matter-dominated universe** today. It was only after the universe became matter-dominated that gas could respond to gravitational interactions with other gas and

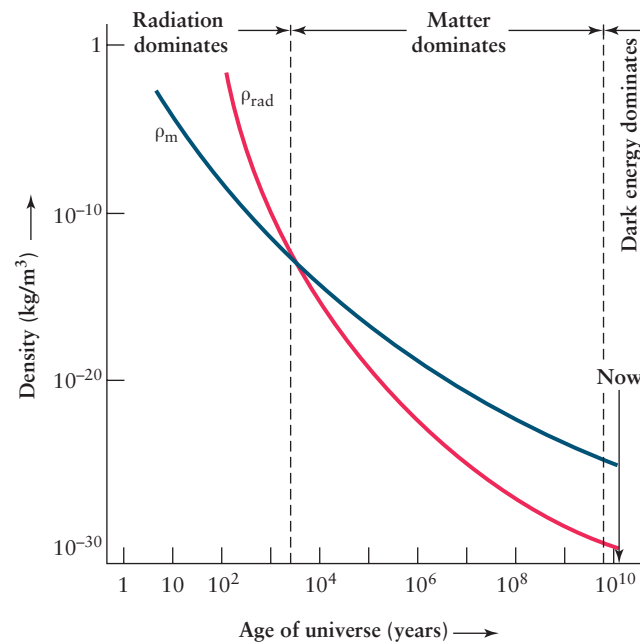


FIGURE 14-13 Evolution of Mass/Energy Density (ρ) For approximately 30,000 years after the Big Bang, the gravitational effects created by photons (ρ_{rad} , shown in red) exceeded the gravitational attraction of all the matter in the universe (ρ_m , shown in blue). This early period is said to have been radiation-dominated. Later, however, continued expansion of the universe caused ρ_{rad} to become less than ρ_m , at which time the universe became matter-dominated.

begin clumping together where the force of gravity was sufficiently strong (an issue we explore further in a moment).

Photons dominated the early universe in another way for about the first 380,000 years. The photons back then could travel only short distances before they collided with particles. These interactions prevented electrons from settling into orbit around nuclei—as soon as electrons did so, photons struck them and kicked them out of orbit, a process called *ionization* (see Section 3-16). During that epoch, the universe was completely filled with a shimmering expanse of high-energy photons colliding vigorously with nuclei and electrons. This state of matter, in which nuclei and electrons are separated from each other, is called a *plasma*. In such an environment you would see a glow from local photons, but because of their vigorous interactions with matter you would not be able to see photons from even a few meters or yards away (Figure 14-14a). The term **primordial fireball** describes the universe during this time.

By around 380,000 years, the universe was so large that the cosmological redshift had stretched photon wavelengths (see Figure 14-1) and thereby decreased photons' energies to the point where they could no longer ionize hydrogen. This stage is called **decoupling**, meaning that the massive particles and the photons

Focus Question 14-8

How, if at all, would the universe be different if antiparticles had exceeded particles in the symmetry breaking described here?

Focus Question 14-9

We say that “most” of the helium in the universe was formed in primordial nucleosynthesis. Where was/is the rest of it formed?

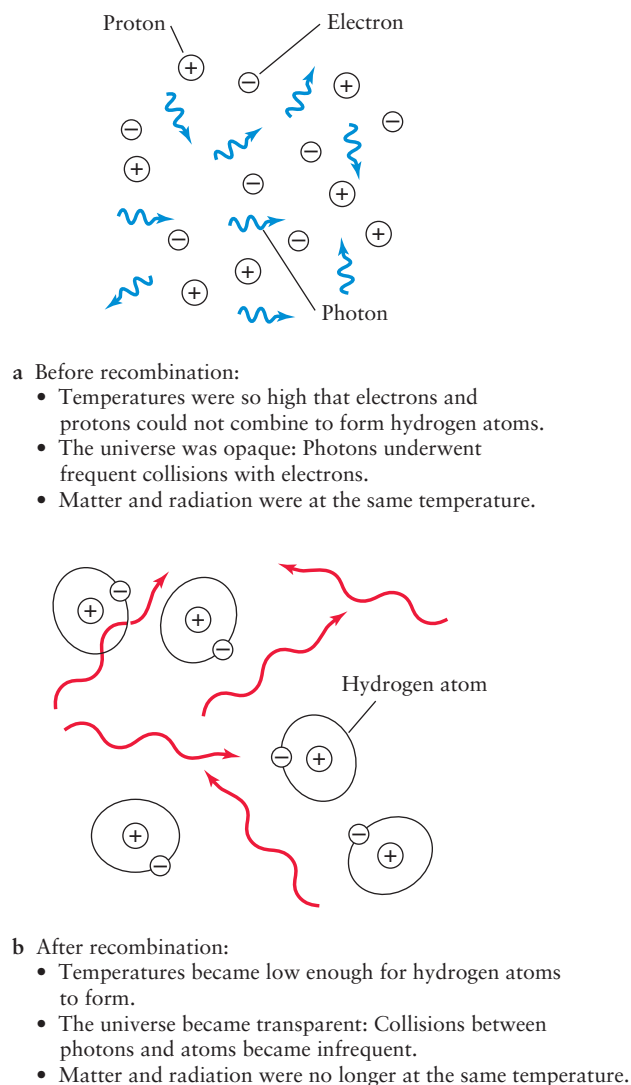


FIGURE 14-14 Era of Recombination (a) Before recombination, the energies of photons in the cosmic background were high enough to prevent protons and electrons from forming hydrogen atoms. (b) As soon as the energy of the background radiation became too low to ionize hydrogen, neutral atoms came into existence.

were effectively disconnected from each other. Electrons began orbiting nuclei, thereby creating neutral atoms (Figure 14-14b). The distance between atoms became so large that most primordial photons now passed by them, rather than colliding with them. Without these collisions to scatter photons, the universe ceased being hazy. It became clear, as it is today, enabling electromagnetic radiation to travel long distances unimpeded and, important from an astronomer's point of view, enabling us to see far into the universe today.

This dramatic period, when the universe transformed from being opaque to transparent, took about 50,000 years to complete and is referred to as the **era of recombination**, referring to nuclei and electrons combin-

ing. There are many instances, in space, on Earth, and in the laboratory, where electrons are stripped from atoms. After such ionization occurs, electrons often recombine with nuclei to recreate neutral atoms. However, the gases created early in the evolution of the universe were never neutral, so the “re” in recombination is misleading. Because the universe was opaque in its first 380,000 years, we cannot see any further into the past than the era of recombination with telescopes that detect electromagnetic radiation of any wavelength.

We can use Wien's law to calculate that the cosmic background radiation had a temperature of 3000 K at the era of recombination. Such a blackbody has a peak microwave wavelength of 0.001 mm (4.0×10^{-5} in). The temperature history of the universe is graphed in Figure 14-8.

Since the primordial gamma-ray photons formed, the expansion of the universe has been increasing their wavelengths, so today we see these same photons as the cosmic microwave background. The universe is now about 1000 times larger than it was at decoupling. Although the universe is dominated by matter today, the number of photons left over from the Big Bang is still immense. From the physics of blackbody radiation, astronomers calculate that there are now 400 million cosmic background microwave photons in every cubic meter of space. In contrast, if all visible matter in the universe were uniformly spread throughout space, there would be roughly one hydrogen atom in every 3 cubic meters (4 cubic yards) of space. In other words, photons continue to outnumber atoms by more than a billion to one. The microwave background contains the most ancient photons we expect ever to be able to observe.

THE STRUCTURE OF THE UNIVERSE

The Big Bang theory explains the evolution of space-time, matter, and energy since the Planck era. It accounts for the hydrogen, most of the helium, and some of the lithium that exist today. But how do we explain the existence of superclusters of galaxies, clusters of galaxies, and individual galaxies?

14-11 Galaxies formed from huge clouds of primordial gas

Had the universe been perfectly uniform until the era of recombination (that is, everywhere perfectly homogeneous and isotropic), it would have been impossible for stars, galaxies, and larger groupings of matter to form, pulled together as they were by differences in mass density and, hence, gravitational attraction. Quantum theory is believed to have played a vital role during the inflationary epoch in creating the clumpiness that we observe today as superclusters of galaxies. In 1927,

the British physicist Paul A. M. Dirac concluded that the laws of nature allow pairs of particles to spontaneously appear, provided that they are identical to each other except for having opposite electric charges, and that they annihilate each other and disappear within a very short period of time, typically 10^{-21} seconds. As discussed in Section 12-18, these pairs are called *virtual particles*.

Normally, *quantum fluctuations* (the spontaneous formation of pairs of virtual particles) occur on the size scales of pairs of particles. However, the quantum fluctuations on microscopic scales that were occurring during the inflationary period were stretched by the inflation until they were large enough to affect the spacetime around them. Those particles created tiny regions of slightly higher temperature, density, and pressure in the inflating universe, like little firecrackers going off all over the place. Also like firecrackers, those events sent out sound waves that traveled through the rapidly expanding universe. These sound waves are believed to have led to the formation of superclusters of galaxies. Here is how that may have taken place.

Whereas light waves, like ocean waves, change amplitude perpendicular to the direction they travel, sound waves change the density of the material through which they move in the direction of their motion. In other words, sound waves are waves that compress and rarefy (make thinner) the medium through which they travel. The rate at which these compressions and rarefactions in the air enter your ears each second, for example, determines the pitches of the sounds you hear. As quantum fluctuations stretched out in the early universe, they compressed the surrounding medium, sending out spherical sound waves, which traveled as long as the gas in the young universe could support them. The speed of these sound waves was impressive by everyday standards. Sound in our atmosphere travels at about 350 m/s (1100 ft/s). The sound waves generated in the inflating universe traveled at 1.75×10^5 m/s (5.75×10^5 ft/s).

The sound waves created at the end of inflation in the early universe traveled out through the plasma of the universe until decoupling, 380,000 years later. At that time, when photons no longer scattered quickly off the gas in the universe, the waves stopped moving.

Because those sound waves were compression and rarefactions of the gas in the young universe, when they stopped traveling, they established shells of higher-density gas (compressions) and interior regions of lower density (rarefactions). The slightly denser and slightly more rarefied regions of the universe that existed at decoupling were also slightly warmer and slightly cooler, respectively, than the average gas in the cosmos. Although those differences in temperature were incredibly small, they have been detected by careful analysis of the data from the COBE and WMAP satellites, which ob-

served the cosmic microwave background temperature at the time of decoupling. Their data (Figure 14-15a) show that these variations in temperature were only about 1 part in 100,000, implying that the average density of matter in the early universe varied by a similar amount. Such tiny differences in density were nevertheless sufficient to start the process of supercluster, cluster, and galaxy formation.

Astrophysicists have plotted the angular sizes of the peaks (warm regions) created by the acoustic (sound) waves, as shown in Figure 14-15b. Their sizes are clustered around 1.2° in angle (Figure 14-15b). In other words, the bubbles of sound generated by the quantum fluctuations spread out to that angular size, as seen from Earth some 13.8 billion light-years away. Allowing for the expansion of the universe since decoupling, those bubbles are about 960 million light-years in diameter today. This size is consistent with predictions of their properties made by the Russian physicist Andrei Sakharov in 1965. Like the sound generated by a violin, there are overtones (higher pitches) in the peaks of the cosmic microwave background. These overtones have different amplitudes. Their relative amplitudes tell us the relative amounts of visible matter and dark matter in the universe, from which we conclude that there is about 5 times as much dark matter as there is visible matter.

We can now outline the scenario for how matter clumped in the early universe. When the universe was radiation-dominated, the particles in it were moving too fast to be pulled together and form structures, like stars, galaxies, clusters, or superclusters. During that time, sound waves from quantum fluctuations created growing shells of higher-density gases, surrounding regions of lower-density gases. Following decoupling, those shells stopped growing. The chemical composition of the early universe was 76% hydrogen and 24% helium, with trace amounts of lithium gas. It took time for the gas to cool sufficiently to start forming stars, a period called the **dark ages**. (Despite the name of that era, the hydrogen in the dark ages was emitting 21-cm photons, as discussed in Section 13-3. These photons have now been redshifted by the Hubble flow to tens of meters (or yards) in length, which is where astronomers are looking for them.) Gravitational attraction in the gases on the shells ultimately led to the formation of superclusters, clusters of galaxies, and galaxies.

The voids seen throughout the cosmos (see Section 13-14) are the relatively empty spaces surrounded by shells of superclusters. The diameters of some of the voids have been measured. They are all 1 billion light-years or smaller, consistent with the 960 million light-year diameters of the shells on which superclusters formed. That not all voids are the same size comes from the fact that many shells of sound crossed each other before decoupling.

The first generation of stars is called *Population III* stars. Except for a trace of lithium, they were complete-

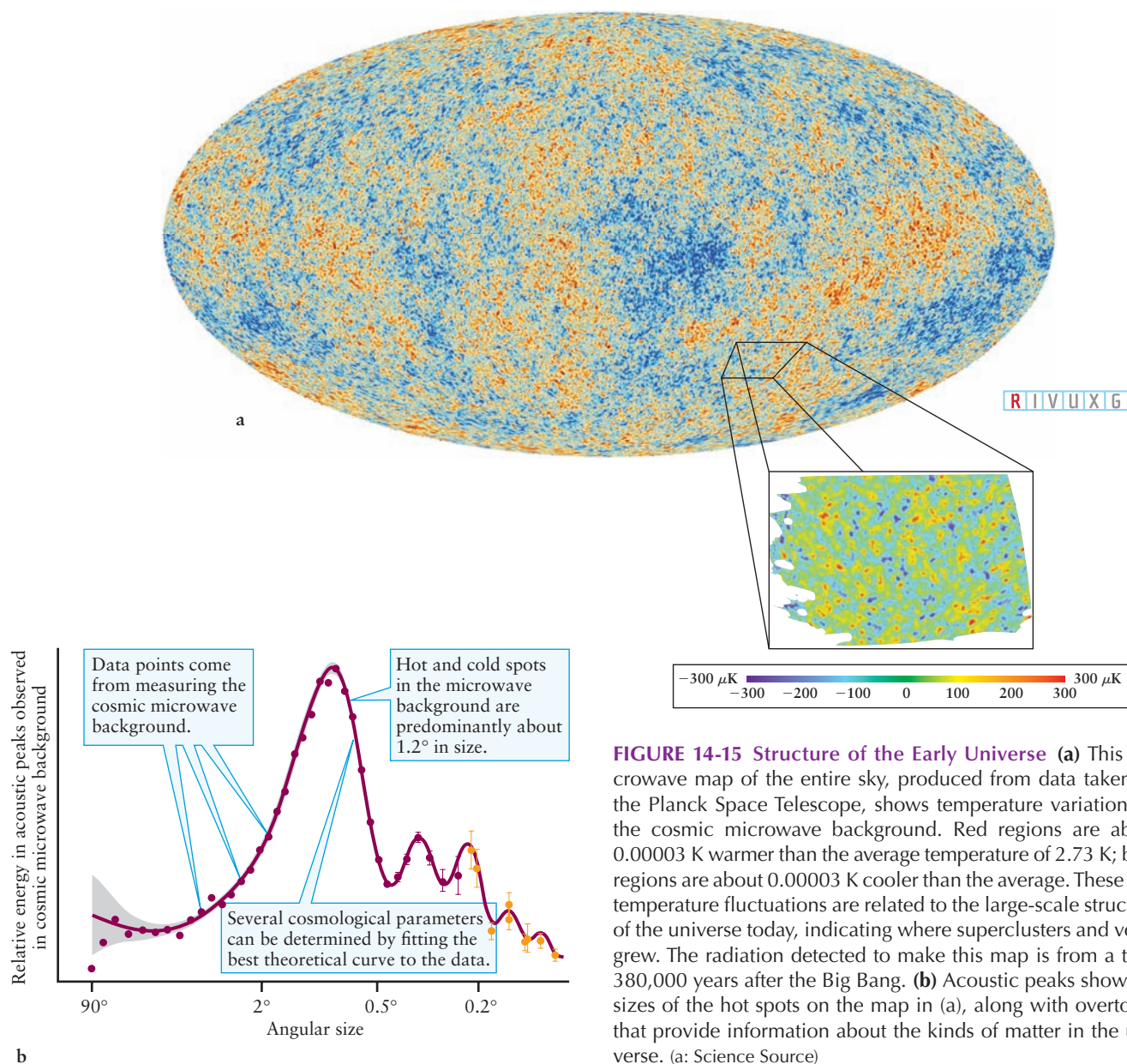


FIGURE 14-15 Structure of the Early Universe (a) This microwave map of the entire sky, produced from data taken by the Planck Space Telescope, shows temperature variations in the cosmic microwave background. Red regions are about 0.00003 K warmer than the average temperature of 2.73 K; blue regions are about 0.00003 K cooler than the average. These tiny temperature fluctuations are related to the large-scale structure of the universe today, indicating where superclusters and voids grew. The radiation detected to make this map is from a time 380,000 years after the Big Bang. (b) Acoustic peaks show the sizes of the hot spots on the map in (a), along with overtones that provide information about the kinds of matter in the universe. (a: Science Source)

ly devoid of metals. Computer simulations indicate that many of these stars had masses in the range 100–500 $M_{\odot}$. These behemoths were a million times or more as bright as the Sun, and they lived for no more than a few million years. Stars with between 100 and 250 $M_{\odot}$ then exploded, while more massive stars collapsed and formed black holes without exploding. The stars in this early cohort that did explode provided the first metals in the universe (other than the primordial lithium). During the first billion years, it appears that clumps of gas that contained millions or billions of solar masses also collapsed to create supermassive black holes. These black holes would have attracted other matter into orbit around them and thereby served as the seeds

for the formation and growth of some galaxies. Recent observations reveal that massive and supermassive black holes typically grow until they have about 0.2% of the mass of a disk galaxy's central bulge.

If, as is now believed, some of the earliest stars had masses in the range of stars found in our Galaxy, it is possible that the low-mass (red dwarf) ones are still on the main sequence. In that case, some of these stars are still orbiting in galaxies like the Milky Way. Because they are Population III stars, astronomers are looking for them through their spectroscopic signature of having virtually no metal. At least one such star has been located. Analogously, astronomers have discovered two intergalactic clouds of gas that, to the precision of the observations,

are composed just of hydrogen, helium, and a trace of lithium. They are remnants of the Big Bang that never became parts of galaxies, nor did they ever undergo star formation, which would have added metals to them.

By observing remote galaxies, astronomers have discovered that most galaxies initially emitted energy we associate with quasars and other active galaxies. This implies that most galaxies have supermassive black holes at their centers. The earliest quasars dating back to when the universe was only 900 million years old provide information about when the first stars formed. Those quasars have spectral lines created by iron and magnesium. In order for those elements to be in those quasars, the iron and magnesium had to be created in stars that formed 700 million years earlier (then exploded and put the metals into the interstellar medium). These metals were then attracted by the black hole. This scenario means that the first stars existed within 200 million years of the Big Bang (Figure 14-16a), thus defining the end of the dark ages.

By 500 million years after the Big Bang, galaxies were beginning to coalesce (Figure 14-16b). We also have

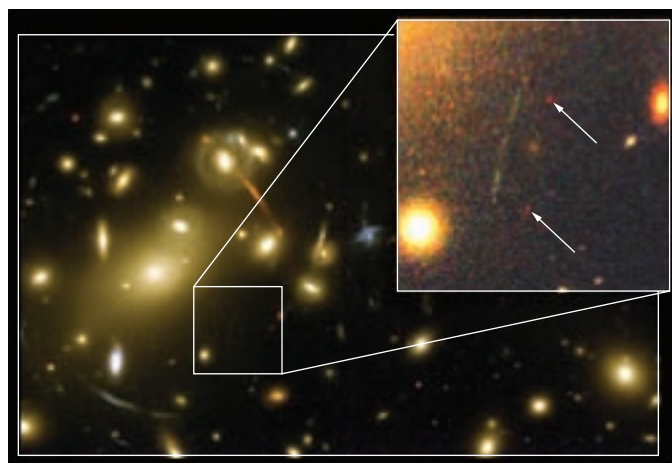
observational evidence that galaxies were in clusters within 2 billion years of the Big Bang (Figure 14-16c).

From about 600 million years to about 2.6 billion years after it formed, the universe underwent heavy star formation. Galaxies grew in size during this period, some continuously, others suddenly and quickly. Star formation decreased as the amount of interstellar hydrogen diminished. One observed galaxy, with about 8 times the mass in stars as the Milky Way has, reached its full star-forming potential when the universe was only about 800 million years old. About 2.6 billion years after the Big Bang there were so many quasars emitting so much energy that much of the gas in the universe became too hot to form new stars, as discussed in Section 13-22. For 500 million years, the star formation rate in galaxies back then decreased. As many quasars became quiescent, the star formation reignited and continued at very high levels until about 6 billion years after the Big Bang.

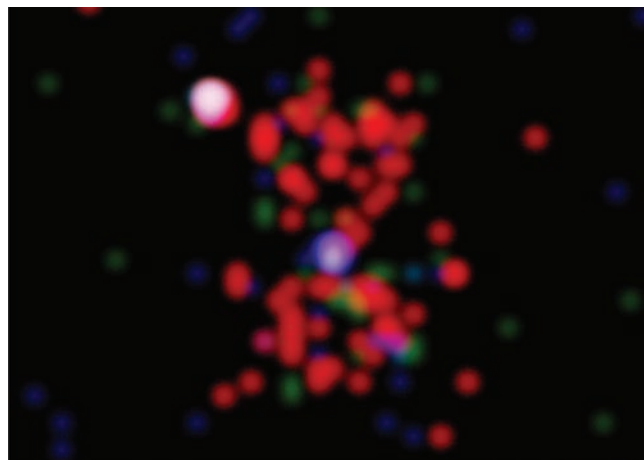
Observations also reveal that galaxies were bluer and brighter in the past than they are today. These changes in color and brightness suggest a high abundance of young,



a Early bursts of star formation



b Arrows indicate galaxies beginning to form 13.4 billion years ago

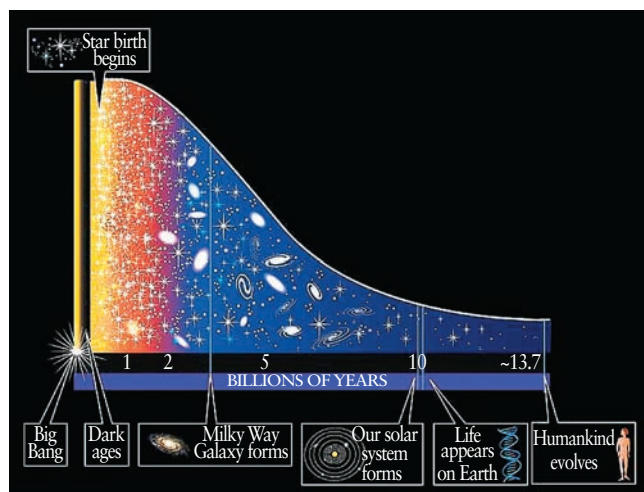


c Massive cluster of galaxies formed 11.2 billion years ago

R I V U X G

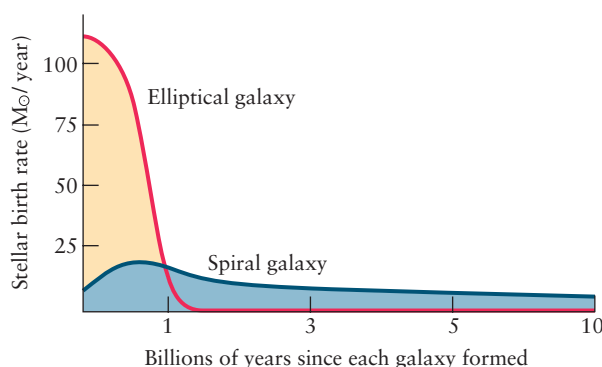
FIGURE 14-16 Galaxies Forming by Combining Smaller Units (a) This image indicates how astronomers visualize the burst of star formation that occurred within a few hundred million years after the Big Bang. The arcs and irregular circles represent interstellar gas illuminated by supernovae. (b) Using the Hubble and Keck telescopes, astronomers discovered two groups of stars (arrows) 13.4 billion light-years away that are believed to be protogalaxies, from which full-grown galaxies form. These protogalaxies were discovered because they were enlarged by the gravitational lensing of an intervening cluster of galaxies. (c) The Chandra X-ray Observatory imaged gravitationally bound gas around the distant galaxy 3C 294. The X-ray emission from this gas is the signature of an extremely massive cluster of galaxies, in this case, at a distance of about 11.2 billion light-years from us. (a: Adolf Schaller, STScI/NASA/K. Lanzetta, SUNY; b: NASA, ESA, A. Fruchter and the ERO Team (STScI, ST-ECF); c: NASA/IoA/A. Fabian et al.)

R I V U X G



a

FIGURE 14-17 Stellar Birth Rates (a) This figure shows that star formation started quickly in the life of the universe and has been tapering off ever since. (b) Most of the stars in an



b

elliptical galaxies are created in a brief burst of star formation when the galaxy is very young. In spiral galaxies, stars form at a more leisurely pace that extends over billions of years.

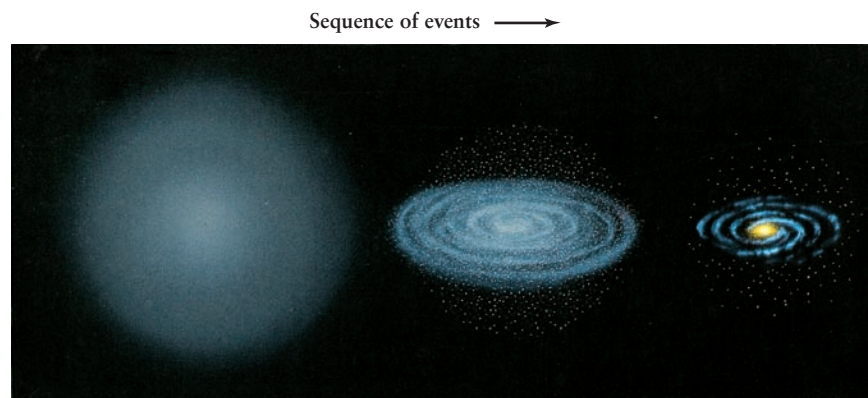
bright, hot, massive stars in newly formed galaxies (Figure 14-17a). As galaxies age, these blue O and B stars become supergiants and eventually die. Therefore, galaxies grow somewhat redder and dimmer. This is especially true of those elliptical galaxies that formed early on. They appear to form nearly all of their stars in one vigorous burst of activity that lasts for about a billion years (Figure 14-17b), after which star formation diminishes drastically and their massive stars evolve and explode so that today we see in them primarily red, intermediate-mass stars. Astronomers say that such elliptical galaxies are “red and dead.”

In contrast, spiral galaxies have been forming stars for at least the past 10 billion years, although at a gradually decreasing rate. There is still plenty of interstellar hydrogen in the disks of spiral galaxies, like our Milky Way, to fuel star formation today. That is why O and B stars still highlight their spiral arms. Figure 14-17b compares the rates at which spiral and elliptical galaxies form stars.

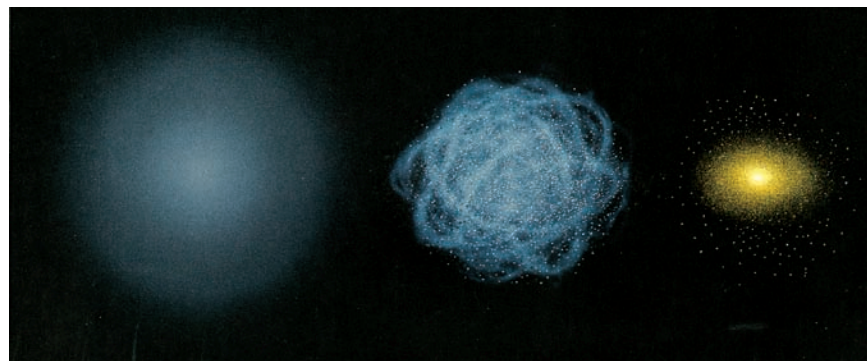
14-12 Star formation activity determines a galaxy's initial structure

Imagine a developing galaxy, called a *protogalaxy*, forming from a cloud of gas. Theory proposes that the rate of star formation determines whether this protogalaxy becomes a spiral or an elliptical galaxy. If stars form slowly enough, then the gas surrounding them has plenty of time to settle by collision with other infalling gas into a flattened disk, just like the early solar system. Star formation continues because the protogalactic disk contains an ample supply of hydrogen, and a spiral or lenticular (disk-shaped but without spiral arms) galaxy is created. If, however, the initial stellar birthrate is high in the protogalaxy,

the theory predicts that virtually all pregalactic gas is used up in the creation of stars before a disk can form. In this case, an elliptical galaxy is created. Figure 14-18 depicts these contrasting sequences of events.



a Formation of a disk galaxy



b Formation of an elliptical galaxy

FIGURE 14-18 Creation of Spiral and Elliptical Galaxies A galaxy begins as a huge cloud of primordial gas that collapses gravitationally. (a) If the rate of star birth is low, then much of the gas collapses to form a disk, and a spiral galaxy is created. (b) If the rate of star birth is high, then the gas is converted into stars before a disk can form, resulting in an elliptical galaxy.

One intriguing nearby galaxy, NGC 1277, apparently didn't play by the normal rules of galactic evolution. It is a lenticular galaxy: It has a disk, but no spiral arms, with a 17 billion $M_{\odot}$ supermassive black hole in its nucleus. Observations suggest that it formed very rapidly

Focus Question 14-10

If the first elements were hydrogen, helium, and lithium, where did Population II stars in globular clusters and elsewhere get their other metals?

(over a 100 million year period) early in the life of the universe and has not changed since then. Normally, such massive black holes take much longer to form. Other lenticular galaxies evolve due to gravitational interactions with neighboring galaxies, but not NGC 1277. Understanding both the

stable lenticular structure and the rapidly formed, extremely massive central black hole are works in progress.

Not all galaxies maintain their initial structure over the evolution of the universe. As we discussed in Section 13-14, some galaxies collide, and these collisions can

change a galaxy's structure from spiral, for example, to elliptical. Indeed, astronomers who use extremely long exposures, called Hubble Deep Field images (for example, see the inset on Figure 14-11), have observed elliptical galaxies during the first few billion

years of the universe's existence that are far less uniform in color than are closer ellipticals. They observed blue stars in the young ellipticals consistent with the merger of spirals and a resulting burst of star formation, as well as lots of lower-mass (yellow and red) stars. Our understanding of galactic formation and evolution is far from complete.

Determining the location and nature of the dark matter in the universe (see Chapter 13) is still of paramount importance in understanding the cosmos. The observable stars, gas, and dust in a galaxy or cluster of galaxies account for only about 10% to 20% of each object's mass. (Recall that this observable matter does not have enough mass to hold galaxies or clusters of galaxies together.) We have very little idea of what the composition of the remaining 80% to 90% of each galaxy or cluster of galaxies is. However, some progress in understanding the distribution of this dark matter is being made. In 2002, astronomers who map large numbers of galaxies and use computer simulations of the effects of dark matter determined that on the scales of clusters of galaxies, the locations of the galaxies often coincide with the concentrations of dark matter, while voids between clusters coincide with voids of dark matter. It is likely that the gravitational force of the dark matter caused gas to concentrate in the same regions, thereby stimulating formation of supermassive black holes and galaxies. In 2007, using images of galaxies that were focused by a gravitational lens of dark matter, astronomers were able to plot the locations of this dark matter (Figure 14-19).

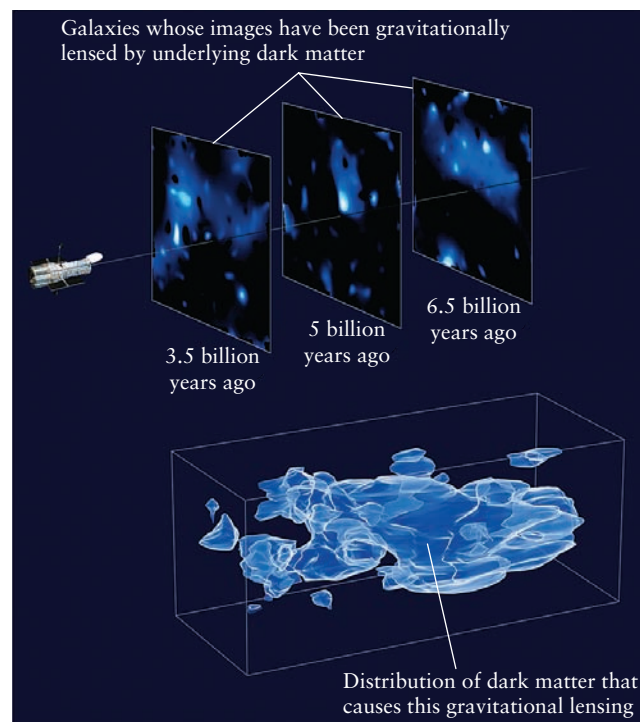


FIGURE 14-19 Mapping Dark Matter (Top) The Hubble Space Telescope observed that galaxies in the same direction, but at different distances from Earth, undergo different amounts of gravitational lensing. **(Bottom)** Much of this effect is due to dark matter. By subtracting out the lensing effects of intervening galaxies, the distorted shapes of the galaxies at various distances enable astronomers to determine the distribution of dark matter. (NASA, ESA, and R. Massey [California Institute of Technology])

THE FATE OF THE UNIVERSE

We now turn to the future. Will the universe last forever? Or will it someday stop expanding and collapse?

14-13 The average density of matter is one factor that determines the future of the universe

As superclusters move apart in the expanding universe, their mutual gravitational attractions act on one another and thereby slow the rate at which they separate. If that were all that mattered in determining the fate of the universe, you could simply locate and add up all of the visible and (presently) dark matter in existence and see if the mutual gravitational force acting between all the components is enough to eventually stop the expansion.

This process is analogous to what engineers do in calculating, for example, how high a cannonball shot upward from the surface of Earth will travel. Earth's gravitational force slows the ball's ascent. If the cannonball's speed upward is less than the escape velocity from Earth's gravity (about 11 km/s or 7 mi/s straight up),

it will fall back to Earth. If the ball's speed exceeds the escape velocity, it will leave Earth and continue outward forever, despite the relentless pull of Earth's gravity. On the boundary between these two scenarios is the situation when the cannonball's speed equals the escape velocity. In that case, the cannonball will just barely escape falling back to Earth, slowing forever, and come to rest at an infinite distance away at an infinite time in the future.

By analogy with the cannonball fired from Earth, it would seem that if the universe is expanding too slowly to overcome the mutual gravitational attraction of its parts, it should stop expanding and someday collapse. If it is expanding at exactly its escape velocity, it should expand until coming to a stop at an infinite time in the future. Or if it is expanding fast enough, it should slow down but continue to expand forever. However, none of these options is correct!

The laws of physics pertaining to the evolution of the universe, as spelled out in the theory of general relativity and in recent observations of distant objects, reveal that reality is more complex than this simple analogy. Just when astronomers were getting comfortable with the idea that the gravitational force is always attractive, general relativity once again showed that reality ignores our common sense by demonstrating that either gravity has a repulsive component or, equivalently, that there is a fifth force in nature that is repulsive on very large scales.

To understand the fate of the universe, we must consider two large-scale factors: the effect of matter on the shape of the universe and the presence of a universe-wide repulsive force named **dark energy** (not to be confused with *dark matter*). We begin by considering the effect that matter and energy have on the shape of the universe.

14-14 The overall shape of spacetime affects the future of the universe

During the 1920s, Alexandre Friedmann in Russia, Georges Lemaître in Belgium, Willem de Sitter in the Netherlands, and Einstein himself applied the theory of general relativity to the expanding universe. General relativity predicts that the presence of matter curves the fabric of spacetime, as we saw in Chapters 12 and 13, in the form of gravitational lensing and the distortion of spacetime around black holes. Similarly, the presence of energy also curves spacetime—recall that matter and energy are related by $E = mc^2$.

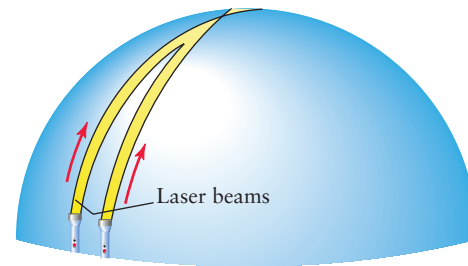
Only three possibilities exist for the overall shape of the universe. These possibilities are determined by the amount of mass and energy the universe contains and how fast it is expanding. For example, imagine shining two powerful laser beams out into space. Suppose that we can align these two beams so that they are perfectly parallel as they leave Earth. Suppose, further, that nothing gets in the way of these two beams. We follow them across the spacetime whose shape we wish to determine. The light beams will

begin to diverge due to the expanding universe carrying them apart. This effect happens regardless of any properties of the spacetime, and, because we are not interested in the effect of expansion right now, we compensate for it (that is, we ignore it) in what follows. The three possibilities for the paths of the laser beams, due to the actual curvature of the universe, are the following:

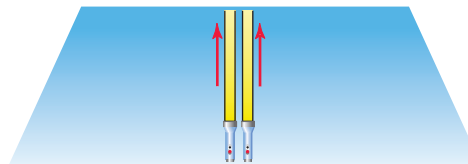
Focus Question 14-12

What, if any, objects have we given enough velocity to escape forever from Earth?

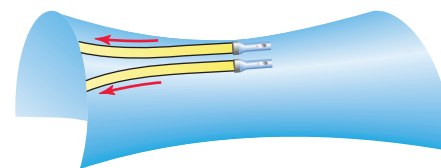
1. Two beams of light, starting out as parallel, gradually get closer and closer together as they move across the universe, eventually intersecting at some enormous distance from Earth (**Figure 14-20a**). In this case, we say that space is *positively curved*. Analogously, the lines of constant longitude on Earth's surface are parallel at the equator but intersect at the poles. Thus, if the universe has this effect on light, the shape of the universe is like the surface of a sphere: Space is spherical and the universe has positive curvature. This shape would happen if the



a Parallel light beams converge



b Parallel light beams remain parallel



c Parallel light beams diverge

FIGURE 14-20 Possible Geometries of the Universe The shape of space (represented here as two-dimensional for ease of visualization) is determined by the matter and energy contained in the universe. The curvature is (a) positive, (b) zero, or (c) negative, depending on whether the average matter and energy density throughout space is greater than, equal to, or less than a critical value. The lines on each curve are initially parallel. They converge, remain parallel, or diverge, depending on the curvature of space.

universe had such high mass and energy densities that the mass and energy literally curved space back in on itself. In the absence of any outward-pushing force, such as would be supplied by a sufficiently large cosmological constant (see Section 14-1), such a universe does not have enough energy to keep expanding forever. It would someday stop expanding and thereafter collapse.

It is easiest to understand such a universe by visualizing a two-dimensional version of it as being the surface of a balloon on which you must remain. In such a universe, you could, in principle, keep going in a straight line (which is a curve on the balloon's surface) and eventually end up back where you started. Like the surface of a balloon, a three-dimensional universe with positive curvature has no outside or center and is said to be a **closed universe**.

2. Two initially parallel beams of light remain parallel regardless of how far they travel (Figure 14-20b). In this case, space is not curved. That is, space is flat and it can extend without limit. This is the structure that meets our commonsense belief.

3. Two initially parallel beams of light gradually diverge farther and farther apart as they move across the universe (Figure 14-20c). This is what would happen if the energy of the Big Bang was sufficiently great to ensure that the universe is going to expand forever. In

this case, the universe has negative curvature. A horse's saddle is a good example of a negatively curved or *hyperbolic* surface. Initially parallel lines drawn on a saddle always diverge. Thus, in a negatively curved universe, we would describe space as hyperbolic. A hyperbolic universe extends without limit and is called an **open universe**. Both the flat and the hyperbolic universes are open and will grow larger forever.

Most astronomers have come to believe that the universe is flat. The reason for this stems from the observations that the universe is homogeneous and isotropic. The only mechanism we have at present to explain these properties of the matter and energy distribution in space is inflation, as discussed earlier. However, the equations predict that if inflation occurred, it stretched the volume of the universe and the matter and energy in it so much that the universe must be very nearly flat.

Telescopic observations strongly support the belief that the universe is flat. This conclusion comes from examining the sizes of the regions of slightly higher and lower temperature in the cosmic microwave background (see Figure 14-15) and comparing them to the sizes predicted for a flat universe. The observations are consistent with the theoretical variations. Furthermore, if the universe had positive curvature, light from the hot regions of the early universe would be curved and thereby focused, creating bigger, brighter images than are observed. **Figure 14-21**

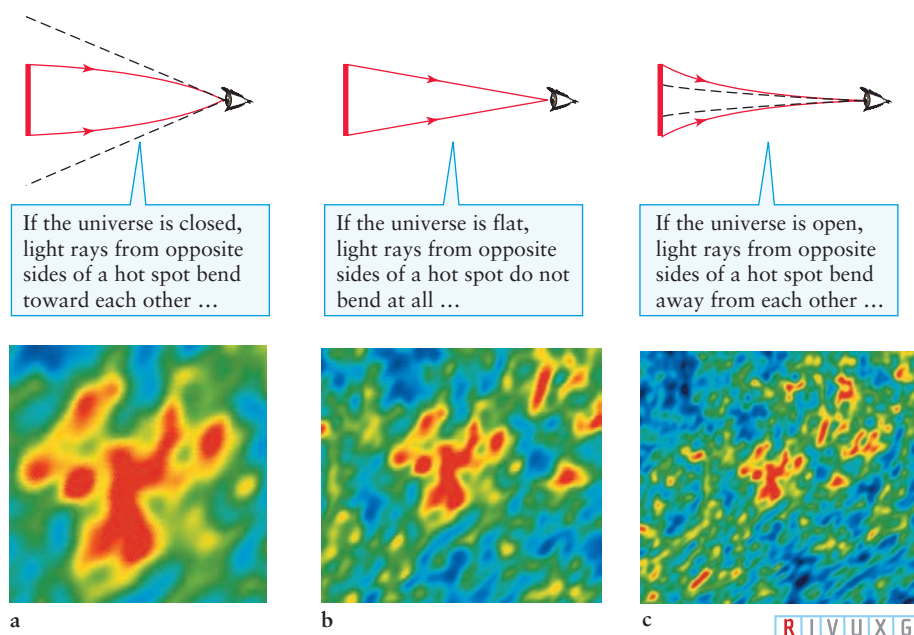


FIGURE 14-21 Cosmic Microwave Background and the Curvature of Space Temperature variations in the early universe appear as “hot spots” in the cosmic microwave background. The apparent sizes of these spots depend on the curvature of space. **(a)** In a closed universe with positive curvature, light rays from opposite sides of a hot spot bend toward each

other. Hence, the hot spot appears larger than it actually is (dashed lines). **(b)** The light rays do not bend in a flat universe. **(c)** In an open universe, light rays bend apart. The dashed lines show that a hot spot would appear smaller than its actual size. (a-c: NASA/WMAP Science Team)

summarizes the effects of space curvature on observations of the cosmic microwave background.

14-15 Dark energy is causing the universe to accelerate outward

Until the past few years, there was a major inconsistency between the distribution of observed matter and energy in the universe and the flatness of space. Just as there is not enough visible matter to account for galaxies and clusters of galaxies that remain as bound systems, there is not enough observed matter and energy to account for a flat universe. Visible matter accounts for only 5% of the required mass. The cosmic microwave background photons add only 0.005% of the required gravitational effects needed for flatness, and calculations of the mass necessary to keep galaxies and clusters bound, combined with gravitational lensing by dark matter of distant galaxies and quasars, reveal that the dark matter known to exist accounts for only about 27% of the required mass. Therefore, the universe has only about 32% of the required mass and energy necessary to make it flat. Allowing for the errors that still exist in all of these observations, the possible range of mass and energy from all matter and photons in the universe is still between only 25% and 30% of that required for flatness. Yet flat it certainly appears to be.

By the mid-1990s, combined evidence of the microwave background and large-scale structure had forced astronomers to the conclusion that in order for the universe to be flat, the remaining energy must be a form of a dark energy with the remarkable feature that it causes the universe to accelerate outward. This energy is considered “dark” because we do not yet know what it is or where it originates.

Corroborating evidence for the existence of such dark energy was found in the recent observations of Type Ia supernovae in extremely distant galaxies (Figure 14-22). The light curves (see Section 12-2) for Type Ia supernovae (explosions of white dwarfs in binary star systems) are very well known. Studies show that these supernovae in nearby galaxies behave similarly to those observed in the Milky Way, regardless of their environment. Because they are so bright, such supernovae are excellent standard candles for determining distances to objects billions of light-years away. Assuming that supernovae in distant galaxies behave similarly to those in our Galaxy (an assumption that is now undergoing careful scrutiny), observations first made in 1998 revealed that the supernovae in distant galaxies appear dimmer than they would if the universe had been continually decelerating or even expanding at a constant speed. In other words, the universe is now expanding faster than it was, so the distant galaxies are farther away than they otherwise would have been if the expansion were slowing or continuing at a constant rate. The universe is *accelerating*

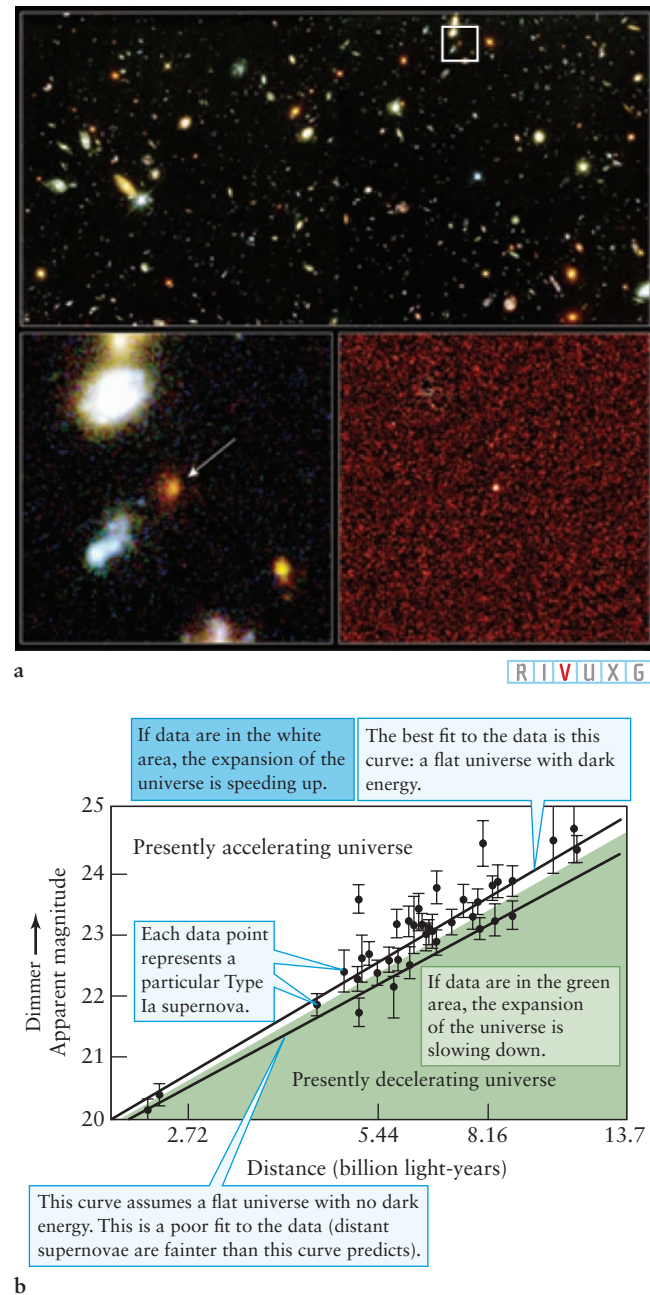


FIGURE 14-22 Dimmer Distant Supernova (a) These Hubble Space Telescope images show the galaxy in which the Type Ia supernova SN 1997ff occurred. This supernova, more than 10 billion light-years away, was dimmer than expected, indicating that the distance to it is greater than the distance it would have if the universe had been continually slowing down since the Big Bang. This supports the notion that an outward (cosmological) force is acting over vast distances in the universe. The arrow on the left inset shows the galaxy in which the supernova was discovered. The bright spot on the right inset shows the supernova by subtracting the constant light emitted by all the other nearby objects. (b) The distances and brightnesses of many very distant supernovae are plotted on this diagram. The locations of the most distant supernovae in the upper region strongly indicate that the universe has been accelerating outward for the past 6 billion years. (a: Adam Riess, Space Telescope Science Institute/NASA)

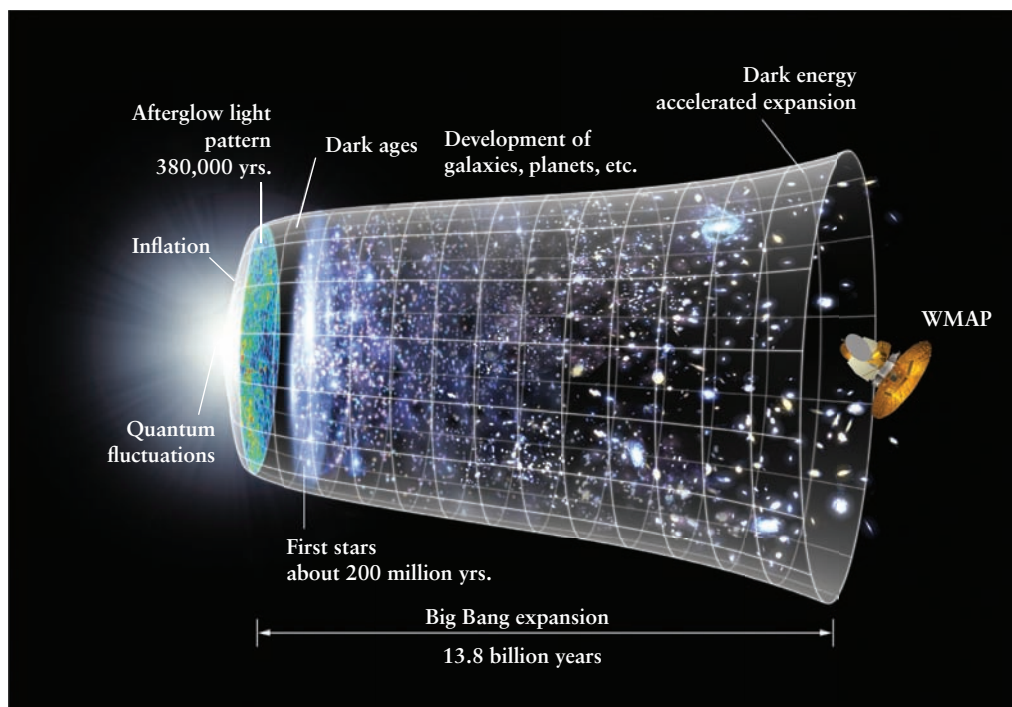


FIGURE 14-23 Big Picture of the Evolution of the Universe This figure shows our current thinking about the evolution of star and galaxy formation in the early universe, as well as the present-day acceleration of the universe's expansion. (NASA/WMAP Science Team)

outward. **Figure 14-23** summarizes the major stages of cosmic evolution discussed in this chapter.

There must be some kind of repulsive force acting to increase the rate at which superclusters separate today. Astronomers hypothesize that this is due to some kind of dark energy that has a repulsive gravitational effect, as mentioned earlier. Confirmation that dark energy exists came in 2003, when astronomers observed light from very distant galaxies passing through clusters of galaxies closer to us on its way toward Earth. As the distant light moved toward a cluster on its way to us, that light was gravitationally blueshifted, meaning the light gained energy as it was pulled toward the cluster. If there were no dark energy in a flat universe, then, as that light left the vicinity of the cluster, it would have been redshifted (lost energy) by exactly the same amount and come to us with the same wavelengths it would have had if it had never passed through the cluster.

However, if dark energy exists, then the universe is accelerating outward (expanding outward faster and faster). So the universe would have expanded more during the time that the light was leaving the cluster than when that light was first moving toward the cluster. The universe's acceleration helps carry the departing photon away from the cluster, so the photon does not have to use as much energy leaving the cluster as it gained traveling toward it. Under those conditions, the light would have gained more energy from the cluster's gravitational attraction while traveling toward the cluster than the light would have lost to that gravitational attraction as it was moving away from the cluster. In other words, the

amount of blueshift that the light underwent falling into the cluster would be greater than the amount of redshift that it underwent leaving the vicinity of the cluster. This was precisely what astronomers discovered: The light from distant galaxies undergoes a net blueshift when it passes through clusters of galaxies on its way to us—the universe is accelerating outward.

Cosmological constant There are presently two viable theoretical explanations for dark energy. Let us first consider what would happen if the cosmological constant that Einstein introduced and then rejected actually did exist. As Einstein had intended, it would provide a repulsive force to nature. The energy associated with that force would appear in the vacuum of space. The cosmological constant form of vacuum energy has the property of contributing a constant repulsive gravitational force that competes with the normal attractive gravitational force from matter and radiation that slows down the expansion. Which gravitational force wins depends on whether there is more vacuum energy or more matter and radiation.

As the universe expands, the average density of matter and energy decreases. In other words, on average there is less matter and energy in each cubic meter or yard of space every second than there was the second before. However, the density of energy created by the cosmological constant is unchanged—as the universe expands, the amount of vacuum energy in each cubic meter of space remains constant. If the universe is still expanding when the vacuum energy exceeds the matter

and energy per unit volume, then the repulsive gravitational force would dominate and force the universe to start accelerating outward, as is observed.

The concern that some astronomers express about the cosmological constant is that the repulsive force it creates has just the right strength to have allowed the universe to slow down for several billion years and then to slowly cause it to accelerate outward, as seen today. To explain this apparent coincidence would require fine-tuning the ratio of the vacuum energy to the matter and radiation to a remarkably high degree of precision. Scientists do not yet have a good reason why the repulsion created by a cosmological constant is not, say, a hundred times greater, in which case structure throughout the universe, such as stars and galaxies, would not have formed yet, or a billion times less, in which case the repulsive effects of the cosmological constant would be negligible today.

Quintessence Despite these concerns, observations of distant Type Ia supernovae suggest that the outward force acting on the universe is nearly constant, consistent with the dark energy being created by the physics we characterized with the cosmological constant. However, these observations do not yet rule out the major competing theory of dark energy, called **quintessence**. Scientists are exploring a variety of mathematical descriptions of quintessence. Quintessence differs from the vacuum energy associated with the cosmological constant in that the energy of quintessence is not constant but changes slowly with the expansion of the universe and can also be changed by the flow of energy and matter through the universe. Because it can change, quintessence need not have been fine-tuned to a strength that makes the universe flat and creates just the slight acceleration we are seeing. Instead, it could have started out with some arbitrary strength and then been adjusted by physical properties of the universe to match the growth of the universe.

As the universe grew, the energy in quintessence would eventually come to dominate the gravitational energy from matter. If this took place a few billion years ago, quintessence would have begun exerting a negative gravitational force and so the universe would have begun accelerating.

5 The unexpected discovery that the universe is accelerating outward completely alters astronomers' perspectives about the fate of the cosmos. This discovery tells us that *the universe will expand forever*. At the same time, it neatly solves the mystery of why the universe is flat, as required by the occurrence of inflation in the early universe. Dark energy provides the remaining 68% of the energy needed to account for the universe being flat. The contributions of all of its major components to the mass and energy in the universe are shown in **Figure 14-24**, with details listed in Appendix L: Mass and Energy Inventory for the Universe.

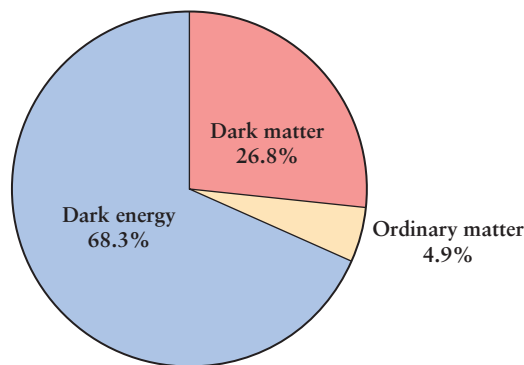


FIGURE 14-24 Percentages of the Major Components of the Universe

We have seen that our present understanding of the laws of nature is incomplete, especially in that our theories do not explain the nature of the universe during the Planck era (or why the universe began in the first place). Driven by these challenges, scientists have begun to develop superstring theories that may resolve these issues. **Discovery 14-1: Superstring Theory and M-Theory** is intended to give you a brief introduction to them and what they now predict.

SUMMARY OF KEY IDEAS

The Big Bang

- Astronomers believe that the universe began as an exceedingly dense cosmic singularity that expanded explosively in an event called the Big Bang. The Hubble law describes the ongoing expansion of the universe and the rate at which superclusters of galaxies move apart.
- The observable universe extends about 13.8 billion light-years in every direction from Earth to what is called the cosmic light horizon. We cannot see any objects that may exist beyond the cosmic light horizon because light from these objects has not had enough time to reach us.
- According to the theory of inflation, early in its existence, the universe expanded very rapidly for a short period, spreading matter that was originally far from our location (and hence at different temperatures and densities) throughout a volume of the universe so large that we cannot yet observe it. The observable universe today is thus a growing volume of space containing matter and radiation that were in close contact with our matter and radiation during the first instant after the Big Bang (and hence at the same temperature, pressure, and density). Inflation explains the isotropic and homogeneous appearance of the universe.

DISCOVERY 14-1

Superstring Theory and M-Theory

To combine gravitation and the other three fundamental forces in nature into one comprehensive *Theory of Everything*, scientists have had to consider a universe that contains more than the four dimensions we know about today (three of space and one of time). The theories that mathematically describe this new formulation of the universe are called *superstring theories*. There are basically five such theories, which allow all of the particles we have been studying—such as quarks, protons, neutrons, electrons, and photons—to exist.

Spacetime in superstring theories has 10 dimensions, of which 6 are everywhere rolled up into such tiny volumes that we cannot detect them directly. The other 4 dimensions are our normal spacetime. Superstring's more general spacetime carries with it properties that allow scientists to combine all four forces in nature into one set of equations.

The difficulty in reconciling quantum mechanics (describing the weak, strong, and electromagnetic forces) and general relativity (describing gravitation) is that the three forces in quantum mechanics are quan-

tized, whereas general relativity is not. In other words, the weak, strong, and electromagnetic forces are transmitted by particles. For example, the quanta of electromagnetism are photons. Gravity, as described by general relativity, is based on a smooth and continuous, rather than quantized, force. Specifically, the distortion of spacetime by matter and energy creates the gravitational force.

Superstring theories begin with a different assumption about all particles and their interactions than do either quantum mechanics or general relativity. The new theories assert that each particle is actually a tiny vibrating string, with different types of particles vibrating at different rates, like different strings on a guitar. Gravitation has its own energy-sharing particle, the graviton, analogous to photons for electromagnetism. The interactions between the strings create all of the properties of matter and energy.

The predictions made by superstring theories begin with the assumption that general relativity is the correct “classical” theory for describing the gravitational interaction between matter and energy. This may seem trivial,

A Brief History of Spacetime, Matter, Energy, and Everything

- Four basic forces—gravity, electromagnetism, the strong nuclear force, and the weak nuclear force—explain the interactions observed in the universe.

- According to current theory, all four forces were identical just after the Big Bang (although our understanding of the nature of matter and energy during this time is extremely limited). At the end of the Planck time (about 10^{-43} seconds after the Big Bang), gravity became a separate force. A short time later, the strong nuclear force became a distinct force. A final separation created the electromagnetic force and the weak nuclear force.

- Observations show that the universe is flat and that the cosmic microwave background is almost perfectly isotropic, resulting from a brief period of very rapid expansion (the inflationary epoch) in the very early universe.

- In its first 30,000 years, the universe was radiation-dominated, during which time photons prevented matter from forming clumps. Then it was matter-dominated, during which time superclusters and smaller clumps of matter formed. Today it is dark energy-dominated. Dark energy of some sort supplies a repulsive gravitational force that causes superclusters to accelerate away from each other.

- During the first 380,000 years of the universe, matter and energy formed an opaque plasma, called the primordial fireball. Cosmic microwave background radiation is the greatly redshifted remnant of the universe as it existed about 380,000 years after the Big Bang.

- About 380,000 years after the Big Bang, spacetime expansion caused the temperature of the universe to fall below 3000 K, allowing protons and electrons to combine and thereby form neutral hydrogen atoms. This period is called the era of recombination. The universe became transparent during the era of recombination, with the photons that existed back then still traveling through space today. In other words, the microwave background radiation is composed of the oldest photons in the universe.

- Clusters of galaxies and individual galaxies formed from pieces of enormous hydrogen and helium clouds, each of which became a separate supercluster of galaxies.

- All of the superclusters and some of the clusters of galaxies within each supercluster are moving away from one another.

- Supermassive black holes appear to have “seeded” the formation of most galaxies.

- During the matter-dominated era, structure formed in the universe. As the universe goes farther into the dark energy-dominated era, the large-scale structure of superclusters of galaxies will fade away.

but, because general relativity today correctly predicts everything in its realm of validity, a more comprehensive (superstring) theory needs to keep that accuracy, or the larger theory is wrong. Some additional predictions of superstring theories include the following:

- The universe cannot have positive curvature (which it does not, as we have seen).
- Some of the clumpy structure that we see as superclusters of galaxies could have been caused by the effects of superstring activity during inflation (with the rest due to the expansion of quantum fluctuations).
- Spacetime may not be entirely smooth. It may have structural defects, like a flawed diamond or ice that has broken into abutting chunks. These defects in spacetime would appear as one-dimensional “cosmic strings” with great density. Their attraction would pull normal matter around them, creating strings of galaxies. These cosmic strings have not yet been found.
- Some of the dark matter may be particles predicted by string theory.

- Most versions of superstring theory include a cosmological constant, but there is no underlying reason yet known for the value of the cosmological constant that may exist today.
- The speed of light is the same for all photons. If different wavelengths of light from the same event arrived at different times, then superstring (and relativity) theories would be wrong.

You may have found the idea of five superstring theories of the universe to be four too many. So do scientists who study string theory. American physicist Ed Witten has shown that in 11 dimensions, all five string theories are equivalent. The one 11-dimensional theory of strings is called M-theory.

Superstring and M-theories are, so far, consistent with observations, but it remains to be seen if they will continue to maintain consistency with future observations and, very importantly, if they will make predictions that can be tested. Without being able to do that, they will remain truly elegant mathematical formalisms, but not science.

The Fate of the Universe

- The average density of matter and dark energy in the universe determines the curvature of space and the ultimate fate of the universe.
- The universe is accelerating outward and it will expand forever.

WHAT DID YOU THINK?

- 1 *What is the universe?* It is all of the matter, energy, and spacetime that will ever be detectable from Earth or that will ever affect us.
- 2 *Did the universe have a beginning?* Yes. It occurred about 13.8 billion years ago, in an event called the Big Bang.
- 3 *Into what is the universe expanding?* Nothing. The Big Bang created space and time (spacetime), as well as all matter and energy in the universe. Spacetime is expanding to accommodate the expansion of the universe.
- 4 *How strong is gravity compared to the other forces in nature?* Gravity is by far the weakest force.
- 5 *Will the universe last forever?* Current observations support the belief that the universe will last forever.

Review Questions

1. Which force in nature is believed to have formed second?
 - a. gravity
 - b. electromagnetic force
 - c. weak force
 - d. strong force
 - e. all formed at the same time
2. Inflation most directly explains which of the following?
 - a. why the universe is homogeneous
 - b. why the universe is expanding
 - c. why the universe is going to last forever
 - d. why the universe has a background temperature
 - e. why the universe has particles
3. What does it mean when astronomers say that we live in an expanding universe?
4. Explain the difference between a Doppler redshift and a cosmological redshift.
5. In what ways are the fate of the universe, the shape of the universe, and the average density of the universe related?
6. Assuming that the universe will expand forever, what will eventually become of the microwave background radiation?

7. What does it mean to say that the universe is dark energy-dominated? When was the universe radiation-dominated? When was it matter-dominated? How did radiation domination show itself?
8. Explain the difference between an electron and a positron.
9. Where do astronomers believe most of the photons in the cosmic microwave background originated?
10. Give examples of the actions or roles of each of the four basic physical forces in the universe.
11. What is the observational evidence for the (a) Big Bang, (b) inflationary epoch, and (c) confinement of quarks?

Advanced Questions

12. Explain why the detection of cosmic microwave background radiation was a major blow to the steady-state theory.
13. Some so-called “creation scientists” claim that the universe came into existence 6000 years ago. What is Hubble’s constant for such a cosmos? Is this a reasonable number? Explain your answers.

Discussion Questions

14. Discuss the implications of the fact that science cannot yet tell us what caused the Big Bang or what, if anything, existed before the Big Bang occurred.
15. Explain why gravitational attraction has dominated the behavior of the universe until recently and why the dark energy determines the fate of the universe.

Got It?

16. Which of the following statements best describes the overall motion in the universe?
 - a. The universe is expanding faster and faster.
 - b. The universe is expanding but more slowly all the time.
 - c. The universe is expanding at a constant rate.
 - d. The universe is collapsing at a constant rate.
 - e. The universe is collapsing faster and faster.
17. What is the weakest force in nature?
18. What do astronomers believe to be the fate of the universe?
19. Has the universe existed forever? If not, when did it come into existence?



CHAPTER 15

Centered on Earth, the red circle shows the region in which astronomers are searching for extraterrestrial intelligence. (NASA/Space Telescope Science Institute)

Astrobiology

WHAT DO YOU THINK?

- 1 Why is water so important to the formation of life?
- 2 What element is uniquely suited to be the foundation of life as we know it, and why?
- 3 How do astronomers search for extraterrestrial intelligence?
- 4 Have astronomers located any extraterrestrial civilizations?
- 5 If advanced alien civilizations exist, is there any way they might know of our existence?

Answers to these questions appear in the text beside the corresponding numbers in the margins and at the end of the chapter.

If, as we said in the first chapter of this book, the starry night sky inspires us to look beyond ourselves, the study of astronomy often motivates us to contemplate the formation of Earth, the nature of the stars, and even the creation of the universe. Perhaps two of the most compelling questions in this realm of nature are how life began here on Earth and whether there is life elsewhere in the universe. Scientists are searching for the origins of life on Earth by doing experiments in hopes of recreating steps in that process. In 2010, for example, a simple artificial cell was fabricated. A key part of these experiments involves trying to determine the physical environment of the young Earth and the materials available on it at the time when life began.

Understanding the origins of life may be helped forward by discovering primitive life elsewhere in our solar system, such as inside Mars, Europa, Ganymede, and, possibly, other moons of the solar system. Astronomers have begun looking for it, developing life-seeking space missions to other worlds, beginning with two that went to Mars in the 1970s (Figure 15-1; see also Section 6-23).

The search for advanced extraterrestrial life has gone in several directions, starting right here on Earth. Conceivably, advanced civilizations from elsewhere

have sent spacecraft to visit our planet. In fact, some people believe that they have seen extraterrestrials or, at least, their spacecraft, commonly called *unidentified flying objects* (UFOs). This discussion has led to the question of whether life on Earth was “seeded” intentionally or accidentally by alien visitors. If Earth was visited, the seed-

ing of it with life by aliens is possible, but Occam’s razor requires us to explore simpler avenues for the beginning of life here.

Are there other advanced civilizations elsewhere in the cosmos? Some people find comfort in the belief that Earth is the only home for life. Others look at the vastness of the universe and the size and diversity of the objects in the cosmos and thereby find reason to expect there may be other worlds suitable for sustaining life. Some scientists are now undertaking high-tech searches

Focus Question 15-1

Describe how you think most scientists would respond to the statement: “Aliens might be here, but we just haven’t seen them yet.”



RIVUXG

FIGURE 15-1 Viking Mars Lander Astronomer and renowned science popularizer Carl Sagan poses by a model of the *Viking* lander. This image was taken in Death Valley, California, where the background creates the feel of a Martian landscape. Sagan was instrumental in choosing some of the experiments flown on the *Viking* spacecraft. (NASA Jet Propulsion Laboratory)

for distant extraterrestrial life, trying to answer this question of whether we are alone in the universe. If we are not alone, perhaps advanced life-forms on other worlds are searching, just as we are, for company in the cosmos.

In this chapter you will discover

- what qualities scientists believe a world must have in order to support life
- why many scientists are open to the possibility that primitive life exists elsewhere in the solar system
- how scientists estimate the number of planets orbiting other stars that could support complex life
- how scientists search for life beyond our own solar system—and the results of those searches
- how we are trying to communicate with advanced extraterrestrial civilizations

15-1 Astrobiology connects the cosmos and the origins of life

The amazing advances in all realms of science that began in the seventeenth century and accelerated to tremendous heights in the past few years have finally enabled scientists to meaningfully explore issues surrounding the beginnings of life on Earth. **Astrobiology** combines all of the natural sciences in an effort to understand the formation, evolution, and future

of life here and throughout the cosmos. As we will explore shortly, efforts to replicate the conditions under which life formed on Earth have been carried out since the 1950s. The discovery of strong evidence that liquid water, essential for life as we know it, exists in several worlds in the solar system has added fuel to the astrobiology effort.

Astrobiology as a unified discipline received a big boost in 1996, when a meteorite from Mars, labeled

ALH84001, was opened to reveal organic-looking structures that *could* have been formed by life on Mars. This possibility sparked the interest of many people, including those who fund NASA and other organizations worldwide. Thousands of scientists now work in this field.

Finding other potential homes for life in the solar system is one facet of astrobiology. Finding homes beyond the solar system is another. The number of known exoplanets has passed 1700. Although most are Jupiterlike (that is, hydrogen-rich bodies), Earthlike planets such as Gliese 581d and HD85512b are now being discovered. Some exoplanets are known to have atmospheres. Some of these planets may have liquid water on their surfaces.

The search for advanced civilizations is also underway. It began on Earth, in the form of people who believe they have seen aliens or, more commonly, their UFO spacecraft. Beliefs are one thing; scientific evidence is something else. Despite all of the alleged UFO sightings and personal encounters that have been reported, no one has produced a single piece of physical evidence to convince scientists that an intelligent, extraterrestrial life-form has *ever* visited Earth. Furthermore, despite all of the wonderful science-fiction books, movies, and TV shows, there is no known way for spacecraft to travel faster than the speed of light. Therefore, unless they come from planets around the few stars within a few light-years of Earth, aliens would have to travel for centuries, or even millenia, to span the void of space between their worlds and ours. After all that travel time, it would seem illogical for space visitors not to communicate with us and eventually land here.

Searches for *extraterrestrial intelligence* (SETI) in space began in 1960, when the American astronomer Frank Drake carried out Project Ozma (named after the ruler of the fictional land of Oz in L. Frank Baum's series of books). Drake used a radio telescope at the National Radio Astronomy Observatory in West Virginia to "listen" for radio signals from civilizations that might have existed around two Sunlike stars: τ (tau) Ceti and ϵ (epsilon) Eridani. He heard nothing.

15-2 The existence of life depends on chemical and physical properties of matter

1 To build a physical foundation for astrobiology, let us begin by considering some of the fundamental requirements for the existence of life as we know it. First is the presence of liquid water. Water, which is neither acidic nor basic, enables many kinds of atoms and molecules dissolved in it to bond to one another or separate from one another in large numbers. While water molecules combine with some atoms and molecules, water also allows many interactions to occur in it without chemical interference. Such diverse activity is essential for the creation and evolution of the complex molecules and systems of molecules that function in living organisms.

Given all of these factors, biologists believe that life on Earth began to develop in primordial bodies of water.

The second requirement is an element that can bond strongly, but not too strongly, with at least three other atoms to allow for complex molecules necessary for life. Atoms able to make only two bonds can form chains (Figure 15-2a), or loops of atoms, but cannot create complex structures (Figure 15-2b) that contain atoms of other elements, as is necessary for making the variety of molecules essential for life.

Five elements can bond covalently (that is, by sharing electrons) with three or more elements: boron (B), carbon (C), nitrogen (N), silicon (Si), and phosphorus (P). Bonding by donating or receiving electrons (called *ionic bonding*) creates bonds that are unsuitable for the formation of the complex, flexible, rapidly changing molecules that life requires. Carbon is unique among the five elements that can make at least three covalent bonds in that its bonds are flexible yet strong. As counterexamples, silicon-silicon bonds come apart under the slightest disturbance; silicon-oxygen bonds create gels and liquids that are very hard to alter (Figure 15-3a); and silicon-oxygen-oxygen bonds are so strong that they create rocks (Figure 15-3b). The other three elements that might serve as the backbone of life—boron, nitrogen, and phosphorus—have similar bonding problems. Despite the enormous range of conditions under

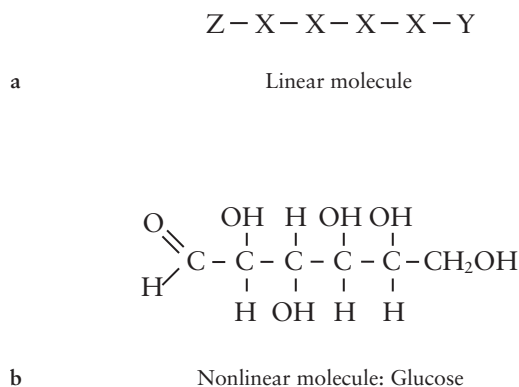


FIGURE 15-2 Creating Complex Molecules (a) Atoms (denoted by capital letters) that can bond strongly to only two other atoms can make linear chains, as depicted here. However, when such atoms bond to atoms that can only make one bond, denoted here as Y and Z, the chain stops. In no case can atoms X, Y, and Z combine to create nonlinear chains other than loops. (b) When an atom, like carbon (C), can share electrons with more than two other atoms (in carbon's case, with four atoms), then the complex, nonlinear chains essential for life can form. Chains of carbon atoms form the backbone of organic molecules. For example, glucose, with carbon, oxygen (O), and hydrogen (H), is a nonlinear molecule that serves as a nutrient for many life-forms; it is a sugar. The lines indicate bonds between atoms.

FIGURE 15-3 Noncarbon Organic Molecules?

When any element other than carbon that can make three or more covalent bonds combines, it makes compounds that are too soft, too hard, too reactive, or too inert to be useful in supporting life. Consider silicon. **(a)** The silicon-oxygen pair that creates the backbone of silicone is too inert to allow such molecules to react rapidly and thereby serve as organic molecules. Furthermore, these bonds produce gel or liquid compounds, as shown. **(b)** When the backbone is silicon-oxygen-oxygen, the bonds are rigid, as in this quartz rock. (a: © 2004 Richard Megna/Fundamental Photographs; b: Roberto Benzi/age fotostock)



a



b

which terrestrial life exists, all of it is based on the unique properties of carbon.

2 Carbon atoms form chemical bonds that can combine to create especially long, complex molecules. These molecules can be further linked in elaborate chains, lattices, and fibers. It is for these reasons that carbon-based compounds, called **organic molecules**, are the stuff of life. Indeed, biologists have already identified more than 6 million carbon compounds.

The variety and stability of living organisms depend on the complex, self-regulating, chemical reactions of organic molecules. Happily, carbon is also among the most abundant elements in the universe. The unique versatility and abundance of carbon imply that extraterrestrial biology would also be based on organic chemistry.

Third, a life-supporting environment must also have a variety of other elements essential for developing life. Other constituents of organic molecules, such as hydrogen, oxygen, nitrogen, and sulfur, are plentiful in our Galaxy. All of the organic molecules necessary to form life may develop naturally in the oceans of Earth and similar planets. In order to expedite the enrichment of oceans with the primordial soup of elements necessary for life, it may be necessary for Earthlike worlds to have large moons, like ours. When they are young and close to their planets (remember from Chapter 6 that our Moon was much closer and is still spiraling away), such moons create powerful tides that can quickly erode the continental shorelines and supply large quantities of minerals to the oceans.

Another way that the formation of life may have been expedited on some worlds is from organic material from space that rained down on them from space debris. This belief is supported by the fact that spectra taken

of comets in our solar system reveal that they contain an assortment of organic compounds. Further evidence for the existence of organic molecules in interplanetary space was collected by, among others, the spacecraft *Giotto* as it passed close to the nucleus of Halley's Comet. In 1997, NASA launched the *Stardust* mission, which rendezvoused with Comet Wild 2 in January 2004, collecting dust and other debris from the comet. The spacecraft returned safely to Earth in January 2006. It revealed the presence of organic molecules that had been released from the comet, supporting the belief that comets may have helped seed compounds useful in the formation of life on Earth.

Evidence of extraterrestrial organic molecules also comes from certain meteorites called *carbonaceous chondrites* (Figure 15-4). Many of the carbonaceous chondrites that have fallen to Earth contain a variety of organic substances. As noted in Section 8-11, carbonaceous chondrites are ancient meteorites dating from the formation of the solar system. If they contained organic molecules while in space, it seems reasonable to conclude that, even from their earliest days, the planets have been continually bombarded with organic compounds.

In interstellar clouds, carbon atoms have combined with other elements to produce an impressive variety of organic compounds. Since the 1960s, radio astronomers have detected telltale microwave emission lines from interstellar clouds that help identify more than 80 of these carbon-based compounds. Examples include ethyl alcohol ($\text{CH}_3\text{CH}_2\text{OH}$), formaldehyde (H_2CO), acetic acid (HCOOCH_3), and acetaldehyde (CH_3CHO). In 2003, scientists detected the amino acid glycine in space. In 2010, molecules with 60 carbon atoms, called buckyballs because they look like structures designed by



FIGURE 15-4 Carbonaceous Chondrite Carbonaceous chondrites are meteorites that date back to the formation of the solar system. This sample is a piece of the Allende meteorite, a large carbonaceous chondrite that fell in Mexico in 1969. (Science Source)

Buckminster Fuller, were observed in a planetary nebula. In 2005, organic compounds called polycyclic aromatic hydrocarbons, or PAHs were observed more than 9 billion light-years away. These were the first observations of complex organic molecules from that far in the past. Because stars and planets condense from interstellar gas and dust enriched in these elements, these compounds can help provide the building blocks of life elsewhere in the Galaxy and in many other galaxies throughout the universe. Indeed, in 2004, astronomers using the Spitzer Space Telescope detected some of the ingredients for life, including methanol, water, and carbon dioxide, in orbit around newly forming stars.

Interstellar space is not the only source of organic material. In a classic experiment performed in 1952, American chemists Stanley Miller and Harold Urey demonstrated that simple chemicals can combine on liquid-water-bearing planets to form organic compounds. Hoping to recreate conditions on primitive Earth and in its atmosphere, as then understood, Miller and Urey mixed hydrogen, ammonia, methane, and water vapor in a closed container. They then subjected their mixture to an electric arc to simulate lightning bolts. At the end of a week, the inside of the container had become coated with a reddish-brown substance rich in organic compounds essential to life (Figure 15-5).

Since that time, geochemists have come to realize that Earth's early atmosphere was more likely to have been a mixture of carbon dioxide, nitrogen, and hydrogen, along with water vapor "outgassed" from volcanoes and deposited by comet impacts. Modern versions of the Miller-Urey experiment have therefore used these common gases, and, once again, organic compounds were produced.

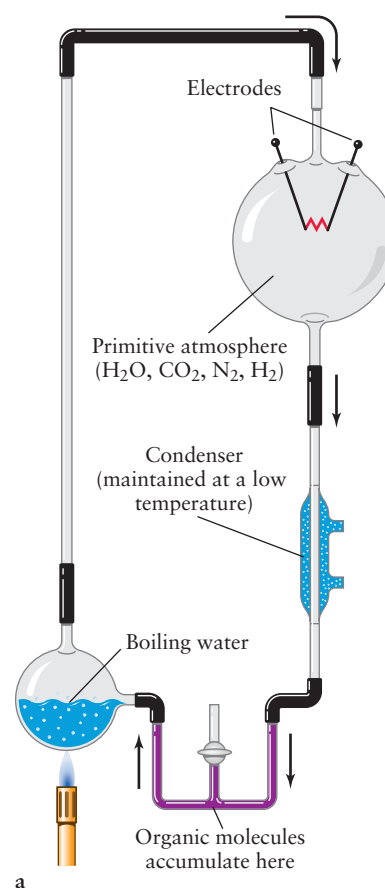


FIGURE 15-5 Miller-Urey Experiment Updated (a) Modern versions of this classic experiment prove that numerous organic compounds important to life can be synthesized from gases that were present in Earth's primordial atmosphere. This experiment supports the hypothesis that life on Earth arose as a result of ordinary chemical reactions. (b) This photograph shows Harold Urey observing his experiment as it was underway. (b: © Bettmann/CORBIS)

Keep in mind that scientists have not yet created life in a test tube. Biologists have yet to figure out, among other things, how simple organic molecules gathered into cells and developed systems for self-replication.

The fourth issue surrounding the existence of life elsewhere relates to physical properties of planets—the environments supporting life-forms must have suitable temperature, radiation, and other factors. The fundamental issue here is that water must be liquid on a world for surface life to evolve there. The region around a star where water can be liquid on a planet is called the star's **habitable zone** (see Chapter 5).

If a planet is too close to a star, the temperature and deadly ultraviolet radiation levels will be too high for life to evolve very far. If the planet is too far away, it will be too cold to have liquid water. Earth is in the Sun's habitable zone, while Venus is not.

If a planet's environment is otherwise hostile—such as having too much radiation, being too windy, or being too seismically active—life may either not be able to spread or may quickly become extinct. If the star is too massive, it will explode before advanced life on any of its planets has a chance to evolve very far. If the star is too small, the planet would have to be very close to it to be warm enough for life, but then tidal forces from the star would lock the planet in synchronous rotation, making most of its surface either too hot (daytime side) or too cold (nighttime side) for life to flourish. Even a world in synchronous rotation with a suitable temperature on the star-facing side is very unlikely to support life, because water in its atmosphere will drift to the night side, permanently freeze, and the star-lit side will eventually become arid. These issues are summarized in **Figure 15-6**.

Focus Question 15-2

Synchronous rotation occurred in what other situation that we explored earlier in this book?

Under the right conditions, moons also have the potential for developing and supporting life. In the habitable zones of stars, sufficiently massive moons (such as Earth-sized terrestrial moons of giant planets in these regions) could sustain liquid water on their surfaces. Even less massive moons could have water below their surfaces kept in the liquid state. This would occur if they orbit their planets in sufficiently elliptical orbits so that changes in the tidal forces from their planets would continually change their shape and thereby warm their insides, as discussed in Chapter 7 for Europa and Ganymede orbiting Jupiter. Such moons would not have to be in the habitable zones of their stars.

Focus Question 15-3

How is water inside worlds like Europa kept liquid?

15-3 Evidence is mounting that life might exist elsewhere in our solar system

Despite all of these potential limitations, scientists have strong evidence that everything does not have to be optimal for life to form. Geologists and biologists have discovered life on Earth in some incredibly challenging

environments, such as on the ocean floor, in a lake far under the Antarctic ice pack, deep inside our planet's crust, and even in hot geothermal vents (**Figure 15-7**). It therefore seems reasonable to believe that life could have originated off Earth under similarly challenging conditions. Scientists consider at least four places in our solar system as possible habitats for life past or present. They are Jupiter's moons Europa, Ganymede, and Callisto, and the planet Mars.

Astronomers have discovered (see Chapter 6) that Mars once had surface liquid water and that it may still have underground bodies of liquid water. We saw in Sections 7-5 through 7-7 that Europa, Ganymede, and Callisto are also likely to have liquid water under their frozen surfaces. These four bodies, then, are candidates for having evolved simple life-forms. Indeed, we saw in Chapter 6 what may be fossilized bacterial life from Mars (see **Figure 6-73b**).

Confirmation that life evolved on any of these worlds would demonstrate that the formation of complex molecules and structures is possible off Earth. However, given the limited chemical resources available on other worlds in our solar system, these locations are probably unsuitable for the evolution of complex life, like ourselves, that is self-aware and able to create advanced civilizations.

15-4 Searches for advanced civilizations try to detect their radio signals

3 We have begun discovering terrestrial exoplanets sufficiently like Earth orbiting stars sufficiently like the Sun so that many astronomers believe it is only a matter of time before we detect the signs of liquid water on some of those worlds. That will set off a flood of effort to try and detect life on them. In the meantime, we are looking for signs of life via radio emission from advanced civilizations on worlds we presently cannot detect. These searches began in the 1950s and 1960s, long before the technology to locate exoplanets was available.

As we have seen, radio waves can travel immense distances without being significantly altered by the interstellar medium. Because they penetrate gas and dust, radio waves are a logical choice for interstellar communication. Even if alien civilizations are not trying to communicate with us, they probably use radio waves in their own technology. But at what frequencies should we look to maximize the odds of detecting alien signals?

We could try random frequencies in the hopes of hearing something—anything. But both the number of directions in which to look and frequencies to check are daunting. We need to find some way to improve the odds. It turns out that some radio wavelengths travel farther without being absorbed by interstellar gas and dust, or without competing with noise from other sources, than do other radio wavelengths. The SETI pioneer Bernard Oliver pointed out that a range of relatively

Intelligent civilizations in our Milky Way Galaxy can evolve only in a certain region called the galactic habitable zone. In that zone, a suitable planet must lie within the planetary habitable zone of its parent star. (Original research by C. H. Lineweaver, Y. Fenner, and B. K. Gibson)

Habitable Zones for Life

Too close to the center of the Milky Way Galaxy:

- The distances between stars are small, so there can be close encounters between stars that would disrupt a planetary system.
- There are also frequent outbursts of potentially lethal radiation from supernovae and from the supermassive black hole at the very center of the Galaxy.

Too far from the center of the Milky Way Galaxy:

- Stars are deficient in elements heavier than hydrogen and helium, so they lack both the materials needed to form Earthlike planets and the chemical substances required for life as we know it.

The Star:

- Must have a mass that is neither too large nor too small.
- If the star's mass is too large, it will use up its hydrogen fuel so rapidly that it will move off the main sequence before life can evolve on any of its planets.
- If the star's mass is too small, the habitable planet would be so close that it would be in synchronous rotation. Water on the star-lit side would vaporize and become locked up as ice on the permanently dark side.

The Neighborhood:

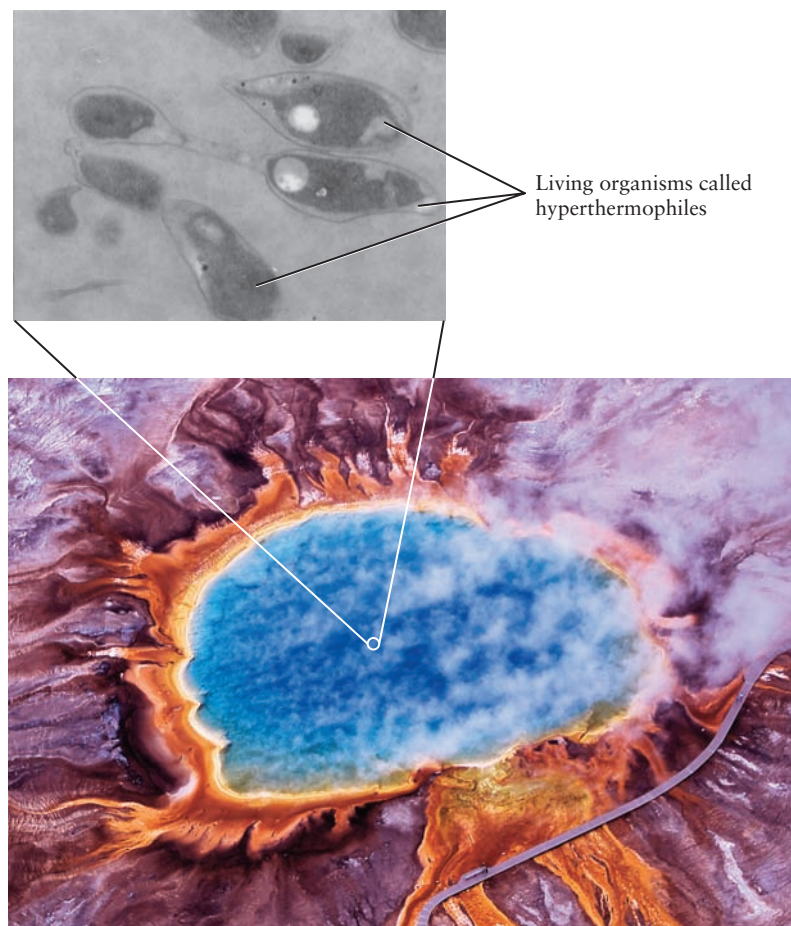
- There needs to be one or more large Jovian planets whose gravitational forces will clear away comets and meteors.

The Planet:

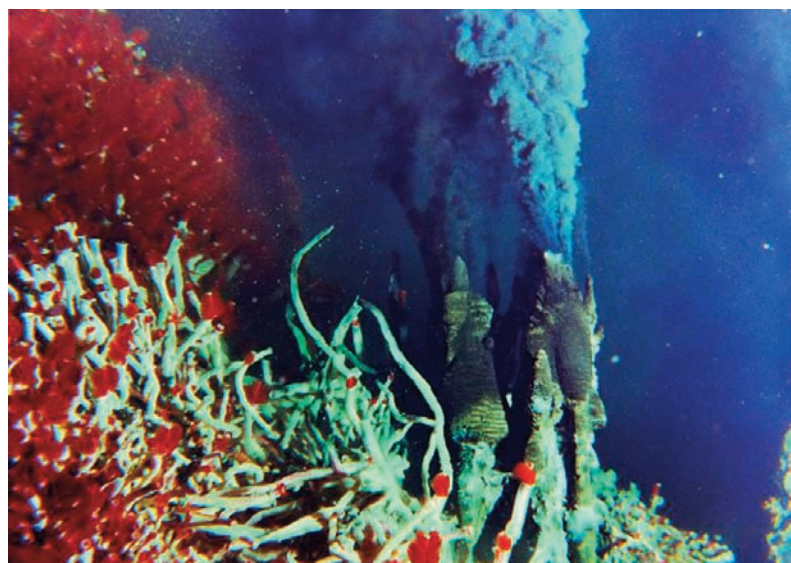
- Must be a terrestrial planet with a solid surface.
- Must have enough mass to provide the gravity needed to retain an atmosphere and oceans.
- Must be at a comfortable distance from the star so that water can be liquid on its surface.
- Must be in a stable, nearly circular orbit. (A highly elliptical orbit would cause excessively large temperature swings as the planet moved toward and away from the star.)

FIGURE 15-6 Zone for Habitable Planets This figure summarizes the locations in the Galaxy and in orbit around stars

where habitable planets might be found. Earth, of course, is in such a location.



a R I V U X G



b R I V U X G

FIGURE 15-7 Hyperthermophiles (a) These microscopic thermophiles (heat-loving organisms) live in water that is between 80°C and 100°C (176°F–212°F). (b) Tube worms (light-green tubes) with hemoglobin-rich red plumes. They reside around black smokers—vents in the ocean bottom that are in the same temperature range as the hot springs shown in (a). These vents are more than 3 km (2 mi) under water. (a: Jim Peaco; July 2001 Yellowstone National Park Image by NPS Photo; inset: NASA; b: V. Tunnicliffe)

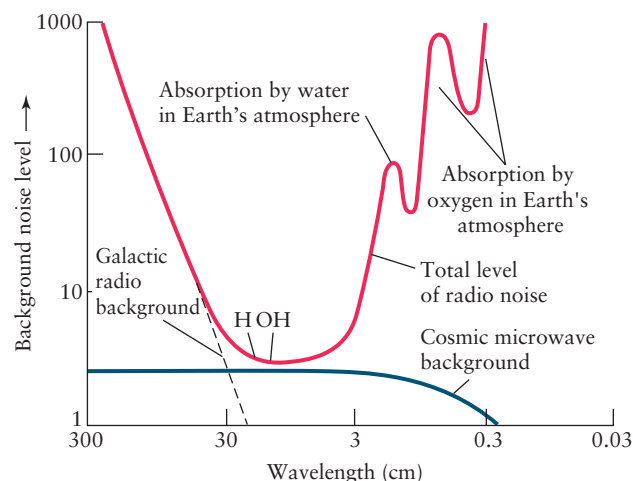


FIGURE 15-8 Water Hole The so-called water hole is a range of radio wavelengths from about 3 to 30 cm that happens to have relatively little cosmic noise. Some scientists suggest that this noise-free region would be well-suited for interstellar communication.

clear frequencies exists in the neighborhood of the microwave emission lines of hydrogen and the hydroxyl radical (OH) (Figure 15-8). This region of the microwave part of the radio spectrum is called the **water hole**, a humorous reference to the H and OH lines being so close together. If extraterrestrial beings are purposefully sending messages into space (to communicate with their spacecraft or with their colonies on other worlds), it seems reasonable that they would choose to transmit on these frequencies.

Even if only a few alien civilizations exist in our neighborhood of the Galaxy, we have the technology to detect radio transmissions from them. The figure at the beginning of this chapter shows the region of the Galaxy from which radio signals similar to the kinds of transmissions we humans make could be detected. Since Project Ozma was carried out, the number of organized searches for signals from extraterrestrials approaches 100. SETI searches use radio telescopes (Figure 15-9) along with sophisticated electronic equipment and powerful computers to conduct both targeted searches and all-sky surveys. Although NASA has created astrobiology centers to study issues regarding life in the solar system and Galaxy, most of the SETI programs in operation today are privately funded. Several programs are actively listening and analyzing enormous volumes of data.

Indeed, because so much data are available to analyze, one SETI project called SETI@home (<http://setiathome.ssl.berkeley.edu>) has enlisted the help of home-computer users. SETI researchers provide actual data from radio telescopes to be analyzed, along with a data-analysis program. While the home computer's



FIGURE 15-9 Radio Telescope Used for SETI The Arecibo Observatory's radio telescope, with a diameter of 305 m (1000 ft), is the largest single-aperture telescope in the world. It is located in Arecibo, Puerto Rico. In the past decade, it has been used in an all-sky survey, along with an antenna located in the Mojave Desert in California, to search for extraterrestrial intelligence. In 1996, an antenna in Canberra, Australia, joined the network. (NAIC—Arecibo Observatory, a facility of the NSF)

screensaver is on, the program runs, the data are analyzed, and the results are sent back to the researchers to be combined with data from other home-computer users. The processed data are replaced with fresh data to be analyzed.

4 Occasionally, a SETI project does detect powerful or unusual signals from space. However, because none of these signals are ever repeated, SETI researchers do not yet believe they have discovered any extraterrestrial civilization.

The detection of a message from an alien civilization would be one of the greatest events in human history. Communicating with aliens could dramatically change the course of civilization on Earth through the sharing of scientific information or an awakening of social or humanistic enlightenment. In only a few years, our industry and social structure could advance centuries into the future. The effects of such change would touch every person on Earth.

15-5 The Drake equation: How many civilizations are likely to exist in the Milky Way?

Just how many planets throughout our Galaxy are likely to harbor complex life? The first person to tackle this question quantitatively was Frank Drake. He proposed that the number of technologically advanced civilizations

in the Galaxy (designated by the letter N) could be estimated by what is now called the **Drake equation**:

$$N = R^* f_p n_e f_i f_c L$$

in which

R^* = rate at which Sunlike stars form in the Galaxy

f_p = fraction of Sunlike stars that have planets

n_e = number of planets per solar-type star system suitable for life

f_i = fraction of those habitable planets on which life actually arises

f_c = fraction of those life-forms that evolve into intelligent species

L = lifetime of that technologically advanced civilization

The Drake equation expresses quantitatively the number of extraterrestrial civilizations as a product of terms, some of which can be estimated from what we know about stars and stellar evolution. For instance, thanks to new evidence for extrasolar planets, astronomers can now determine the first two terms, R^* and f_p , by observation. We should probably exclude stars larger than about $1.5 M_\odot$ because they have main-sequence lifetimes shorter than the time it took for intelligent life to develop on Earth—some 3.8 to 4.0 billion years. If that period is typical of the time needed to evolve higher life-forms, then a massive star would become a giant or even explode as a supernova before self-aware creatures could evolve on any of its planets.

Although low-mass stars have much longer lifetimes, they are less well-suited than stars in the range of $1 M_\odot$ for supporting life on their planets because they are so cool. As noted earlier, only planets very near a low-mass star would be sufficiently warm for water to be a liquid. But, as also mentioned earlier, a planet that close would become tidally coupled to its star and develop synchronous rotation. One side would have continuous daylight, leading to the evaporation of oceans, while the other side would be in perpetual, frigid darkness. The only place that life could survive on such planets is in the narrow ring at the boundary between day and night. Such a small fraction of the land available greatly reduces the likelihood that life would be able to evolve to the complexity necessary for technological civilizations to develop.

As “ideal” life-supporting stars, this leaves main-sequence stars with masses near those of the Sun. These

Focus Question 15-4

If we detect incidental, rather than directed, signals from alien civilizations, what information might they contain?

stars are of spectral types between F5 and M0. Based on statistical studies of star formation in the Milky Way, astronomers calculate that roughly one of these Sunlike stars forms in the Galaxy each year, yielding a value of $R^* = 1$ per year.

We learned in Sections 4-2 through 4-7 that the planets in our solar system formed in conjunction with the birth of the Sun. We have also seen evidence that similar processes of planetary formation may be commonplace around isolated stars. Many astronomers, therefore, assign f_p a value of 1, meaning they believe it is likely that most Sunlike stars have planets.

The chances that a planetary system has an Earthlike world are not yet known, although on-going observations of exoplanets should soon reveal this number. Let us hypothesize that one in five solar-type stars is orbited by a habitable planet, making $n_e = 0.2$.

Unfortunately, the rest of the factors in the Drake equation are very hypothetical. From what we know about the evolution of life on Earth, we might assume that, given appropriate conditions, the development of life is a certainty, which would make $f_i = 1$. This value is, of course, an area of intense interest to biologists. For the sake of argument, we might also assume that evolution naturally leads to the development of intelligence (a conjecture that is hotly debated) and also make $f_i = 1$. It is anyone's guess as to whether these intelligent extraterrestrial beings would attempt communication with other civilizations in the Galaxy, but, if we assume they all would, f_c would also be put at 1.

The last variable, L , the longevity of technological civilization, is the most uncertain of all—it cannot be tested! Looking at our own example, we see a planet whose atmosphere and oceans are increasingly polluted, potentially destroying the food chain. When we add in how close we have come to destroying ourselves with weapons of mass destruction, it may be that we humans are among the lucky few technological civilizations to squeak through its first years. In other words, L may be as short as 100

years. Let's be a little more optimistic and assume that on average $L = 500$ years. Putting all of these numbers together, we arrive at

$$N = 1 \times 1 \times 0.2 \times 1 \times 1 \times 1 \times 500 = 100$$

Therefore, out of the hundreds of billions of stars in the Galaxy, there may be only 100 civilizations technologically advanced enough to communicate with us. Of course, the numbers used here are just estimates. A wide range of values has been proposed for the terms in the Drake equation, and these various numbers produce vastly different estimates of N , the total number.

Increasing the average lifetime of advanced civilizations significantly increases N . Some scientists argue that there is exactly one advanced civilization in the Galaxy and that we are it. Others speculate that there may be tens of millions of planets inhabited by intelligent creatures. Although we do not know yet, science enables us to home in on the number.

15-6 Humans have been sending signals into space for more than a century

5 We have been doing more than just listening passively for voices from the cosmos. Astronomers have intentionally broadcast well-focused signals through radio telescopes toward star systems likely to harbor advanced life, including a 1974 “hello” message aimed at the globular cluster M13 (Figure 15-10a) and a 2008 broadcast of the Beatles’ song “Across the Universe.” Unintentional broadcasts into space have occurred for longer than that. Since 1895, when the first radio transmission was made by the Italian engineer Guglielmo Marconi, we have been sending messages about ourselves into space. The spherical region within some 120 light-years of the solar system is now filled with signals from radio and television shows. If other advanced civilizations fall within that sphere, they could very well be listening to radio programs or watching television shows from decades past, trying to make sense of our species.

Four spacecraft also carry information about the human race. Launched in 1972 and 1973, respectively, the *Pioneer 10* and *Pioneer 11* spacecraft carry plaques about us (Figure 15-10b). They are now making their way into interstellar space. In 1977, the two *Voyager* spacecraft were sent toward the outer solar system. They carry old-fashioned phonograph records encoded with human images and voices. Because space is so vast and the spacecraft are so small, the likelihood of their being discovered is truly remote. And yet millions, perhaps billions, of years from now, one of the spacecraft just might reach another race of intelligent creatures. From its path and the information on board, those creatures might be able to determine where the *Pioneer* and *Voyager* spacecraft came from. If their travel budgets are sufficiently large, they might even decide to come and visit us. At the least, they could send us radio messages. Will our descendants be here to receive them?

SUMMARY OF KEY IDEAS

- The chemical building blocks of life exist throughout the Milky Way Galaxy.
- Organic molecules and water have been discovered in interstellar clouds, in some meteorites, in comets, and in newly forming star and planet systems.

Focus Question 15-5

With what two numbers used in the Drake equation calculation presented previously do you disagree most? Using your numbers, how many civilizations do you estimate exist?

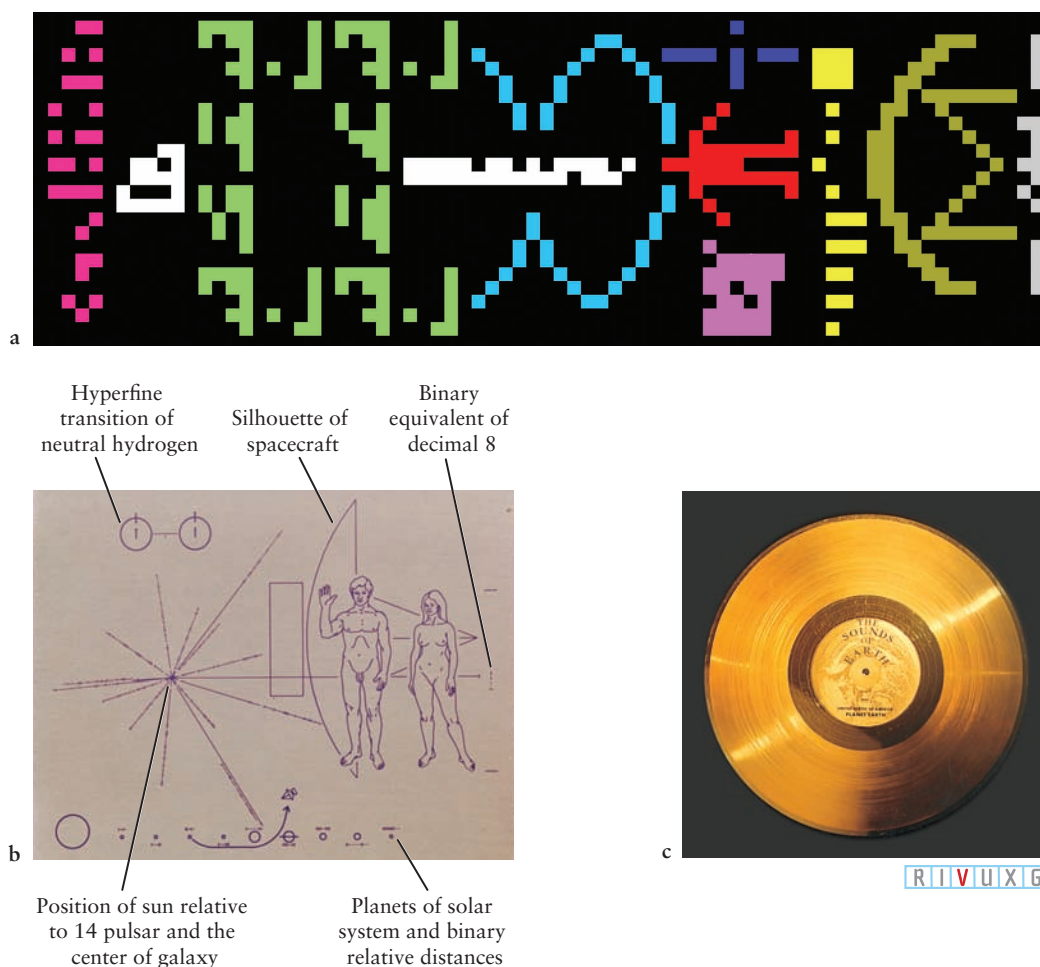


FIGURE 15-10 Memorabilia in Space (a) Humans have beamed radio signals into space, hoping that the message will someday be intercepted by an alien civilization. This is a visual version of the signal sent in 1974 from the Arecibo radio telescope toward the globular cluster M13. The *Pioneer* and *Voyager* spacecraft, now in interstellar space, also carry messages from Earth. (b) The plaques on *Pioneer 10* and *Pioneer 11* provide information about where humans are, what we look like, and some of the science we know. (c) Images

and sounds sent on *Voyager 1* and *Voyager 2* were stored on phonograph records, long before DVDs were even a twinkle in an engineer's eye. There are also instructions for playing the record, which contains information about our biology, our technology, and our knowledge base. Each record also contains the sounds of children's voices. It is remotely possible that another race might someday discover the spacecraft. (a: SPL/Science Source; b, c: NASA)

Focus Question 15-6

What negative consequence could arise from an alien civilization discovering a *Pioneer* spacecraft?

- Astronomers are using radio telescopes to search for signals from other self-aware life in the Galaxy. This effort is called the search for extraterrestrial intelligence, or SETI. SETI is primarily done at frequencies where radio waves pass most easily through the interstellar medium. So far, these searches have not detected any life outside Earth.
- The Drake equation is used to estimate the number of technologically advanced civilizations in the Galaxy whose radio transmissions we might discover. Estimates of this number vary from one to millions.
- Everyday radio and television transmissions from Earth, along with intentional broadcasts into space, may be detected by other life-forms.

WHAT DID YOU THINK?

- 1 *Why is water so important to the formation of life?* Water allows many interactions between atoms and molecules dissolved in it.
- 2 *What element is uniquely suited to be the foundation of life as we know it, and why?* Carbon is the only atom that can bond flexibly (but not too flexibly) with three or more other atoms.
- 3 *How do astronomers search for extraterrestrial intelligence?* They search for radio signals from other advanced civilizations.

4. *Have astronomers located any extraterrestrial civilizations?* No. Extraterrestrial civilizations have not yet been discovered.

5. *If advanced alien civilizations exist, is there any way they might know of our existence?* Yes. They might detect our intentional radio broadcasts into space, our everyday radio and television broadcasts, or, possibly, one of our spacecraft traveling in interstellar space.

Discussion Questions

1. Which stellar spectral type is most likely to have a planet on which advanced life exists?

- | | |
|------|------|
| a. O | d. G |
| b. B | e. M |
| c. A | |

2. Which element serves as the foundation or “backbone” of life?

- | | |
|------------|-------------|
| a. oxygen | d. hydrogen |
| b. carbon | e. nitrogen |
| c. silicon | |

3. What arguments could you make against sending messages via radio or spacecraft into interstellar space?

4. Try using the Drake equation with values that you find reasonable. How many civilizations do you estimate there are in our Galaxy?

5. Discuss the possible biological effects of our probes that visit other life-sustaining worlds.

6. What social effects might probes from other worlds have on us?

7. List some of the pros and cons in the argument that alien spacecraft have visited Earth.

Got It?

8. Have astronomers found concrete evidence for the existence of extraterrestrial life? If so, please describe these discoveries.

9. Have scientists found credible evidence that extraterrestrials have visited Earth? Explain your answer.

10. Has life been created in the laboratory? If so, when and where?

APPENDICES

APPENDIX A: POWERS-OF-TEN NOTATION

Astronomy is a science of extremes. As we examine various cosmic environments, we find an astonishing range of conditions—from the incredibly hot, dense centers of stars to the frigid, near-perfect vacuum of interstellar space. To describe such divergent conditions accurately, we need a wide range of both large and small numbers. Astronomers avoid such confusing terms as “a million billion billion” (1,000,000,000,000,000,000,000,000) by using a standard shorthand system. All the cumbersome zeros that accompany such a large number are consolidated into one term consisting of 10 followed by an *exponent*, which is written as a superscript and called the **power of ten**. The exponent merely indicates how many zeros you would need to write out the long form of the number. Thus,

$$10^0 = 1$$

$$10^1 = 10$$

$$10^2 = 100$$

$$10^3 = 1000$$

$$10^4 = 10,000$$

and so forth. Equivalently, the exponent tells you how many tens must be multiplied together to yield the desired number. For example, ten thousand can be written as 10^4 (“ten to the fourth”) because $10^4 = 10 \times 10 \times 10 \times 10 = 10,000$. Similarly, 273,000,000 can be written as 2.73×10^8 .

In *scientific notation*, numbers are written as a figure between 1 and 10 multiplied by the appropriate power of 10. The distance between Earth and the Sun, for example, can be written as 1.5×10^8 km (or 9.3×10^7 mi). Once you get used to it, you will find this notation more convenient than writing “150,000,000 kilometers” or “one hundred and fifty million kilometers.”

This powers-of-ten system can also be applied to numbers that are less than 1 by using a minus sign in front of the exponent. A negative exponent tells you that the location of the decimal point is as follows:

$$10^0 = 1.0$$

$$10^{-1} = 0.1$$

$$10^{-2} = 0.01$$

$$10^{-3} = 0.001$$

$$10^{-4} = 0.0001$$

and so forth. For example, the diameter of a hydrogen atom is approximately 1.1×10^{-8} cm or (4.3×10^{-9} in). That is more convenient than saying “0.000000011 centimeter,” “11 billionths of a centimeter,” or the equivalent numbers in U.S. customary units. Similarly, 0.000728 equals 7.28×10^{-4} .

Using the powers-of-ten shorthand, one can write large or small numbers like these compactly:

$$3,416,000 = 3.416 \times 10^6$$

$$0.000000807 = 8.07 \times 10^{-7}$$

Because powers-of-ten notation bypasses all the cumbersome zeros, a wide range of circumstances can be numerically described conveniently:

$$\text{one thousand} = 10^3$$

$$\text{one million} = 10^6$$

$$\text{one billion} = 10^9$$

$$\text{one trillion} = 10^{12}$$

and also

$$\text{one thousandth} = 10^{-3} = 0.001$$

$$\text{one millionth} = 10^{-6} = 0.000001$$

$$\text{one billionth} = 10^{-9} = 0.000000001$$

$$\text{one trillionth} = 10^{-12} = 0.000000000001$$

Try these questions: Write 3,141,000,000 and 0.0000000031831 in scientific notation. Write 2.71828 $\times 10^{10}$ and 3.67879×10^{-11} in standard notation. (Answers appear at the end of the book.)

APPENDIX B: TEMPERATURE SCALES

Three temperature scales are in common use. Throughout most of the world, temperatures are expressed in degrees Celsius ($^{\circ}\text{C}$), named in honor of the Swedish astronomer Anders Celsius, who proposed it in 1742. The **Celsius temperature scale** (also known as the “centigrade scale”) is based on the behavior of water, which freezes at 0°C and boils at 100°C at sea level on Earth.

Scientists usually prefer the **Kelvin scale**, named after the British physicist Lord Kelvin (William Thomson), who made many important contributions to our

knowledge about heat and temperature. On the Kelvin temperature scale, water freezes at 273 K and boils at 373 K. Note that we do not use the degree symbol with the Kelvin temperature scale.

Because water must be heated by 100 K or 100°C to go from its freezing point to its boiling point, you can see that the size of a kelvin is the same as the size of a degree Celsius. When considering temperature *changes*, measurements in kelvins and in degrees Celsius lead to the same number.

A temperature expressed in kelvins is always equal to the temperature in degrees Celsius plus 273. Scientists prefer the Kelvin scale because it is closely related to the physical meaning of temperature. All substances are made of atoms, which are very tiny (a typical atom has a diameter of about 10^{-10} m or 3.9×10^{-9} in) and constantly in motion. The temperature of a substance is directly related to the average speed of its atoms. If something is hot, its atoms are moving at high speeds. If a substance is cold, its atoms are moving much more slowly.

The coldest possible temperature is the temperature at which atoms move as slowly as possible (they can never quite stop completely). This minimum possible temperature, called *absolute zero*, is the starting point for the Kelvin scale. Absolute zero is 0 K, or -273°C . Because it is impossible for anything to be colder than 0 K, there are no negative temperatures on the Kelvin scale.

In the United States, many people still use the now outdated Fahrenheit scale, which expresses temperature in degrees Fahrenheit ($^\circ\text{F}$). When the German physicist Gabriel Fahrenheit introduced this scale in the early 1700s, he intended 0°F to represent the coldest temperature then achievable (with a mixture of ice and saltwater) and 100°F to represent the temperature of a healthy human body. On the Fahrenheit scale, water freezes at 32°F and boils at 212°F . Because there are 180 degrees Fahrenheit between the freezing and boiling points of water, a degree Fahrenheit is only $100/180$ ($= 5/9$) the size of the other scales.

The following equation converts from degrees Fahrenheit to degrees Celsius:

$$T_C = 5/9 (T_F - 32)$$

To convert from Celsius to Fahrenheit, a simple rearrangement of terms gives the relationship

$$T_F = 9/5 T_C + 32$$

where T_F is the temperature in degrees Fahrenheit and T_C is the temperature in degrees Celsius.

For example, consider a typical room temperature of 68°F . Using the first equation, we can convert this measurement to the Celsius scale as follows:

$$T_C = 5/9 (68 - 32) = 20^\circ\text{C}$$

To arrive at the Kelvin scale, we simply add 273 degrees to the value in degrees Celsius. Thus, $68^\circ\text{F} = 20^\circ\text{C} = 293$ K. **Figure B-1** displays the relationships among these three temperature scales.

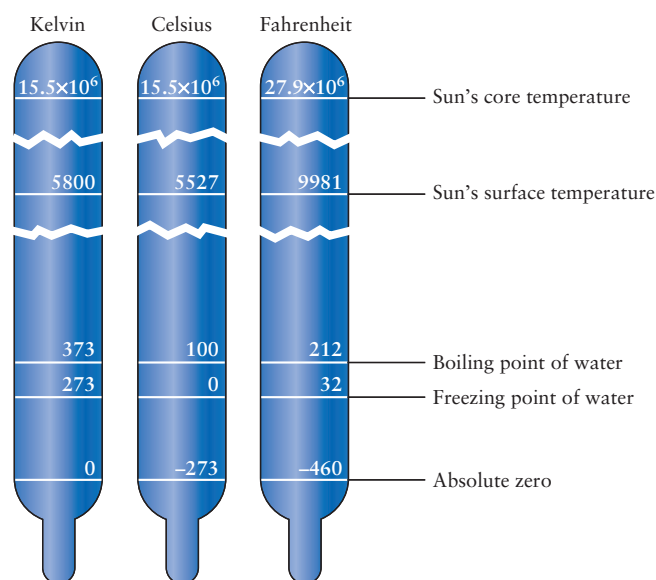


Figure B-1 Temperature Scales Compared

Try these questions: The Sun's surface temperature is about 5800 K. What is its temperature in Celsius and Fahrenheit? The temperature of empty space is about 3 K. What is its temperature in Celsius and Fahrenheit? (Answers appear at the end of the book.)

APPENDIX C THE PLANETS: ORBITAL DATA

Planet	Semimajor axis		Sidereal period		Synodic period (day)	Mean orbital speed (km/s)	Orbital eccentricity	Inclination of orbit to ecliptic (°)
	(AU)	(10 ⁶ km)	(year)	(day)				
Mercury	0.3871	57.9	0.2408	87.97	115.88	47.9	0.206	7.00
Venus	0.7233	108.2	0.6152	224.70	583.92	35.0	0.007	3.39
Earth	1.0000	149.6	1.0000	365.26	—	29.8	0.017	0.00
Mars	1.5237	227.9	1.8808	686.98	779.94	24.1	0.093	1.85
Jupiter	5.2034	778.6	11.862	4,332.6	398.9	13.1	0.048	1.31
Saturn	9.5371	1433.5	29.457	10,759	378.1	9.7	0.054	2.48
Uranus	19.1913	2872.5	84.01	30,685	369.7	6.8	0.047	0.77
Neptune	30.0690	4495.1	164.79	60,189	367.5	5.4	0.009	1.77

APPENDIX D THE PLANETS: PHYSICAL DATA

Planet	Equatorial diameter		Mass		Mean density (kg/m ³)	Rotation period* (days)	Inclination of equator to orbit (°)	Surface gravity (Earth = 1)	Albedo	Escape speed (km/s)
	(km)	(Earth = 1)	(kg)	(Earth = 1)						
Mercury	4,879	0.383	3.302×10^{23}	0.055	5427	58.646	173.4	0.38	0.106	4.3
Venus	12,104	0.949	4.869×10^{24}	0.815	5243	243.02 ^R	177.4	0.91	0.65	10.4
Earth	12,756	1.000	5.974×10^{24}	1.000	5515	0.996	23.45	1.000	0.37	11.2
Mars	6,794	0.533	6.419×10^{23}	0.107	3933	1.025	25.19	0.38	0.15	5.0
Jupiter	142,984	11.209	1.899×10^{27}	317.83	1326	0.413	3.13	2.5	0.52	59.5
Saturn	120,536	9.449	5.685×10^{26}	95.16	687	0.445	26.73	1.1	0.47	35.5
Uranus	51,118	4.007	8.683×10^{25}	14.54	1270	0.717 ^R	97.77	0.91	0.51	21.3
Neptune	49,528	3.883	1.024×10^{26}	17.147	1638	0.671	28.32	1.1	0.4	23.5

* For Jupiter, Saturn, Uranus, and Neptune, the internal rotation period is given. A superscript *R* means that the rotation is retrograde (opposite the planet's orbital motion).

APPENDIX E MAJOR SATELLITES OF THE PLANETS BY MASS

Planet	Satellite	Discoverers of moons	Average distance from center of planet (km)	Orbital (sidereal) period* (days)	Orbital eccentricity	Diameter of satellite (km)	Mass (kg)
EARTH	Moon	—	384,400	27.322	0.0549	3476	7.349×10^{22}
MARS	Phobos	Hall (1877)	9,378	0.319	0.02	$28 \times 23 \times 2$	1.1×10^{16}
	Deimos	Hall (1877)	23,459	1.262	0.00	$16 \times 12 \times 10$	2.4×10^{15}
JUPITER	Io	Galileo (1610)	421,600	1.769	0.00	3643	8.93×10^{22}
	Europa	Galileo (1610)	670,900	3.551	0.01	3138	4.80×10^{22}
	Ganymede	Galileo (1610)	1,070,000	7.155	0.00	5268	1.48×10^{23}
	Callisto	Galileo (1610)	1,883,000	16.689	0.01	4821	1.08×10^{23}
SATURN	Mimas	Herschel (1789)	185,520	0.942	0.02	392	3.8×10^{19}
	Enceladus	Herschel (1789)	238,020	1.370	0.00	444	7.3×10^{19}
	Tethys	Cassini (1684)	294,660	1.888	0.00	1050	6.3×10^{20}
	Dione	Cassini (1684)	377,400	2.737	0.00	1120	1.1×10^{21}
	Rhea	Cassini (1672)	527,040	4.518	0.00	1528	2.3×10^{21}
	Titan	Huygens (1655)	1,221,830	15.945	0.03	5150	1.3×10^{23}
	Iapetus	Cassini (1671)	3,561,300	79.330	0.03	1436	1.6×10^{21}
URANUS	Miranda	Kuiper (1948)	129,872	1.413	0.00	−470	6.6×10^{19}
	Ariel	Lassell (1851)	190,945	2.520	0.00	−1160	1.4×10^{21}
	Umbriel	Lassell (1851)	265,998	4.144	0.01	1169	1.2×10^{21}
	Titania	Herschel (1787)	436,298	8.706	0.00	1578	3.5×10^{21}
	Oberon	Herschel (1787)	583,519	13.463	0.00	1523	3.0×10^{21}
NEPTUNE	Triton	Lassell (1846)	354,760	5.877^R	0.00	2704	2.14×10^{22}
	Nereid	Kuiper (1949)	5,513,400	360.1	0.75	340	2×10^{19}

*A superscript R means that the satellite orbits in a retrograde direction (opposite to the planet’s rotation).

APPENDIX F THE NEAREST STARS

Name*	Parallax (arcsec)	Distance (ly)	Spectral type	Radial velocity** (km/s)	Proper motion (arcsec/year)	Apparent visual magnitude	Absolute visual magnitude	Luminosity (Sun = 1)
Sun			G2 V			−26.7	+4.85	1.00
Proxima Centauri	0.769	4.22	M5.5 V	−22	3.853	+11.09	+15.53	8.2×10^{-4}
Alpha Centauri A	0.747	4.40	G2 V	−25	3.710	−0.01	+4.38	1.77
Alpha Centauri B	0.747	4.40	K0 V	−21	3.724	+1.34	+5.71	0.55
Barnard's Star	0.547	5.94	M4 V	−111	10.358	+9.53	+13.22	3.6×10^{-3}
Wolf 359	0.419	7.80	M6 V	+13	4.696	+13.44	+16.6	3.5×10^{-4}
Lalande 21185	0.393	8.32	M2 V	−84	4.802	+7.47	+10.44	0.023
L 726−8 A	0.374	8.56	M5.5 V	+29	3.368	+12.54	+15.4	9.4×10^{-4}
L 726−8 B	0.374	8.56	M6 V	+32	3.368	+12.99	+15.9	5.6×10^{-4}
Sirius A	0.380	8.61	A1 V	−9	1.339	−1.43	+1.47	26.1
Sirius B	0.380	8.61	white dwarf	−9	1.339	+8.44	+11.34	2.4×10^{-3}
Ross 154	0.337	9.71	M3.5 V	−12	0.666	+10.43	+13.07	4.1×10^{-3}
Ross 248	0.316	10.32	M5.5 V	−78	1.617	+12.29	+14.8	1.5×10^{-3}
Epsilon Eridani	0.310	10.49	K2 V	+17	0.977	+3.73	+6.19	0.40
Lacaille 9352	0.304	10.73	M1.5 V	+10	6.896	+7.34	+9.75	0.051
Ross 128	0.299	10.87	M4 V	−31	1.361	+11.13	+13.51	2.9×10^{-3}
L 789−6	0.294	11.09	M5 V	−60	3.259	+12.33	+14.7	1.3×10^{-3}
61 Cygni A	0.286	11.36	K5 V	−65	5.281	+5.21	+7.49	0.16
61 Cygni B	0.286	11.44	K7 V	−64	5.172	+6.03	+8.31	0.095
Procyon A	0.286	11.40	F5 IV−V	−4	1.259	+0.38	+2.66	7.73
Procyon B	0.286	11.40	white dwarf	−4	1.259	+10.7	+12.98	5.5×10^{-4}
BD +59° 1915 A	0.281	11.61	M3 V	−1	2.238	+8.94	+11.18	0.020
BD +59° 1915 B	0.281	11.61	M3.5 V	+1	2.313	+9.70	+11.97	0.010
Groombridge 34 A	0.281	11.65	M1.5 V	+12	2.918	+8.08	+10.32	0.030
Groombridge 34 B	0.281	11.65	M3.5 V	+11	2.918	+11.06	+13.3	3.1×10^{-3}
Epsilon Indi	0.276	11.82	K5 V	−40	4.704	+4.69	+6.89	0.27
GJ 1111	0.276	11.82	M6.5 V	−5	1.290	+14.78	+16.98	2.7×10^{-4}
Tau Ceti	0.274	11.90	G8 V	−17	1.922	+3.49	+5.68	0.62
GJ 1061	0.272	12.08	M5.5 V	−20	0.826	+13.09	+15.26	1.0×10^{-3}
L 725−32	0.269	12.12	M4.5 V	+28	1.372	+12.10	+14.25	1.7×10^{-3}
BD +05° 1668	0.263	12.40	M3.5 V	+18	3.738	+9.84	+11.94	0.011
Kapteyn's star	0.255	12.79	M1.5 V	+246	8.670	+8.84	+10.87	0.013
Lacaille 8760	0.253	12.89	M0 V	+28	3.455	+6.67	+8.69	0.094
Krüger 60 A	0.248	13.05	M3 V	−33	0.990	+9.79	+11.76	0.010
Krüger 60 B	0.248	13.05	M4 V	−32	0.990	+11.41	+13.38	3.4×10^{-3}

*Stars that are components of binary systems are labeled A and B.

**A positive radial velocity means the star is receding; a negative radial velocity means the star is approaching.

Compiled from the *Hipparcos General Catalogue* and from data reported by the Research Consortium on Nearby Stars. The table lists all known stars within 4.00 parsecs (13.05 light-years).

APPENDIX G THE VISUALLY BRIGHTEST STARS

Name	Designation	Distance (ly)	Spectral type	Radial velocity* (km/s)	Proper motion (arcsec/year)	Apparent visual magnitude	Apparent visual brightness** (Sirius = 1)	Absolute visual magnitude	Luminosity (Sun = 1)
Sirius A	α CMa A	8.61	A1 V	−9	1.339	−1.43	1.000	+1.47	26.1
Canopus	α Car	313	F0 I	+21	0.031	−0.62	0.470	−5.53	1.4×10^4
Arcturus	α Boo	36.7	K2 III	−5	2.279	−0.05	0.278	−0.31	190
Rigel Kentaurus	α Cen A	4.4	G2 V	−25	3.71	−0.01	0.268	+4.38	1.77
Vega	α Lyr	25.3	A0 V	−14	0.035	+0.03	0.258	+0.58	61.9
Capella	α Aur	42.2	G8 III	+30	0.434	+0.08	0.247	−0.48	180
Rigel	β Ori A	773	B8 Ia	+21	0.002	+0.18	0.225	−6.69	7.0×10^5
Procyon	α CMi A	11.4	F5 IV−V	−4	1.259	+0.38	0.184	+2.66	7.73
Achernar	α Eri	144	B3 IV	+19	0.097	+0.45	0.175	−2.77	5250
Betelgeuse	α Ori	427	M2 Iab	+21	0.029	+0.45	0.175	−5.14	4.1×10^4
Hadar	β Cen	525	B1 II	−12	0.042	+0.61	0.151	−5.42	8.6×10^4
Altair	α Aql	16.8	A7 IV−V	−26	0.661	+0.77	0.132	+2.2	11.8
Aldebaran	α Tau A	65.1	K5 III	+54	0.199	+0.87	0.119	−0.63	370
Spica	α Vir	262	B1 V	+1	0.053	+0.98	0.108	−3.55	2.5×10^4
Antares	α Sco A	604	M1 Ib	−3	0.025	+1.06	0.100	−5.28	3.7×10^4
Pollux	β Gem	33.7	K0 III	+3	0.627	+1.16	0.091	+1.09	46.6
Fomalhaut	α PsA	25.1	A3 V	+7	0.368	+1.17	0.090	+1.74	18.9
Deneb	α Cyg	3230	A2 Ia	−5	0.002	+1.25	0.084	−8.73	3.2×10^5
Mimosa	β Cru	353	B0.5 III	+20	0.05	+1.25	0.084	−3.92	3.4×10^4
Regulus	α Leo A	77.5	B7 V	+4	0.249	+1.36	0.076	−0.52	331

Data in this table compiled from the *Hipparcos General Catalogue*.
*A positive radial velocity means the star is receding; a negative radial velocity means the star is approaching.
**This is the ratio of the star's apparent brightness to that of Sirius, the brightest star in the night sky.

Note: Acrux, or α Cru (the brightest star in Crux, the Southern Cross), appears to the naked eye as a star of apparent magnitude +0.87, the same as Aldebaran, but it does not appear in this table because Acrux is actually a binary star system. The blue-white component stars of this binary system have apparent magnitudes of +1.4 and +1.9; thus, they are dimmer than any of the stars listed here.

APPENDIX H THE CONSTELLATIONS

Name	Meaning	R.A.	Dec.	Genitive*	Abbreviation
Andromeda	proper name; princess	1	+40	Andromedae	And
Antlia	air pump	10	−35	Antliae	Ant
Apus	bee	16	−75	Apodis	Aps
Aquarius ^{1,2}	waterman	22	−10	Aquarii	Aqr
Aquila	eagle	20	+15	Aquilae	Aql
Ara	altar	17	−55	Arae	Ara
Aries ²	ram	3	+20	Arietis	Ari
Auriga	charioteer	6	+40	Aurigae	Aur
Boötes	proper name; herdsman, wagoner	15	+30	Boötis	Boo
Caelum	engraving tool	5	−40	Caeli	Cae
Camelopardalis	giraffe	6	+70	Camelopardalis	Cam
Cancer ²	crab	8.5	+15	Cancri	Cnc
Canes Venatici	hunting dogs	13	+40	Canum Venaticorum	CVn
Canis Major	larger dog	7	−20	Canis Majoris	CMA
Canis Minor	smaller dog	8	+5	Canis Minoris	CMi
Capricornus ^{1,2}	water—goat	21	−20	Capricornii	Cap
Carina	keel	9	−60	Carinae	Car
Cassiopeia	proper name; queen	1	+60	Cassiopeiae	Cas
Centaurus	centaur	13	−45	Centauri	Cen
Cepheus	proper name; king	22	+65	Cephei	Cep
Cetus	whale	2	−10	Ceti	Cet
Chamaeleon	chameleon	10	−80	Chamaeleontis	Cha
Circinus	compasses	15	−65	Circini	Cir
Columba	dove	6	−35	Columbae	Col
Coma Berenices	Berenice's hair	13	+20	Comae Berenices	Com
Corona Australis ³	southern crown	19	+40	Coronae Australis	CrA
Corona Borealis ⁴	northern crown	16	+30	Coronae Borealis	CrB
Corvus ⁵	crow, raven	12	−20	Corvi	Crv
Crater	cup	11	−15	Crateris	Crt
Crux ⁶	southern cross	12	−60	Crucis	Cru
Cygnus	swan	21	+40	Cygni	Cyg
Delphinus ¹	dolphin	21	+10	Delphini	Del
Dorado ⁷	swordfish	6	−55	Doradus	Dor
Draco ⁸	dragon	15	+60	Draconis	Dra
Equuleus	little horse	21	+10	Equulei	Equ
Eridanus	proper name; river	4	−30	Eridani	Eri
Fornax	furnace	3	−30	Fornacis	For
Gemini ²	twins	7	+20	Geminorum	Gem

APPENDIX H THE CONSTELLATIONS (Continued)

Name	Meaning	R.A.	Dec.	Genitive*	Abbreviation
Grus	crane	22	−45	Gruis	Gru
Hercules ⁹	proper name; hero	17	+30	Herculis	Her
Horologium	clock	3	−55	Horologii	Hor
Hydra	water serpent	12	−25	Hydrae	Hya
Hydrus	water snake	2	−70	Hydri	Hyi
Indus	Indian	22	−70	Indi	Ind
Lacerta	lizard	22	+45	Lacertae	Lac
Leo ²	lion	11	+15	Leonis	Leo
Leo Minor	smaller lion	10	+35	Leonis Minoris	LMi
Lepus	hare	6	−20	Leporis	Lep
Libra ^{2,10}	scales	15	−15	Librae	Lib
Lupus	wolf	15	−45	Lupi	Lup
Lynx	lynx	8	+45	Lyncis	Lyn
Lyra ⁵	lyre	19	+35	Lyrae	Lyr
Microscopium	microscope	21	−40	Microscopii	Mic
Monoceros	unicorn	7	0	Monocerotis	Mon
Mensa	table	6	−75	Mensae	Men
Musca (Australis) ¹¹	(southern) fly	12	−70	Muscae	Mus
Norma	square	16	−50	Normae	Nor
Octans ¹²	octant	—	−90	Octantis	Oct
Ophiuchus ^{2,13}	serpent-bearer	17	0	Ophiuchi	Oph
Orion	proper name; hunter, giant	6	0	Orionis	Ori
Pavo	peacock	20	−70	Pavonis	Pav
Pegasus	proper name; winged horse	23	+20	Pegasi	Peg
Perseus	proper name; hero	3	+45	Persei	Per
Phoenix	phoenix	1	−50	Phoenicis	Phe
Pictor	easel	6	−55	Pictoris	Pic
Pisces ^{1,2}	fishes	1	+10	Piscium	Psc
Piscis Austrinus ¹	southern fish	22	−30	Piscis Austrini	PsA
Puppis	stern	8	−30	Puppis	Pup
Pyxis	compass	9	−30	Pyxidis	Pyx
Reticulum	net	4	−60	Reticuli	Ret
Sagitta	arrow	20	+20	Sagittae	Sge
Sagittarius ^{2,14}	archer	19	−25	Sagittarii	Sgr
Scorpius ²	scorpion	17	−30	Scorpii	Sco
Sculptor ¹⁵	sculptor's workshop	1	−30	Sculptoris	Scl
Scutum ¹⁶	shield	19	−10	Scuti	Sct
Serpens ¹³	serpent	17	0	Serpentis	Ser

APPENDIX H THE CONSTELLATIONS (Continued)

Name	Meaning	R.A.	Dec.	Genitive*	Abbreviation
Sextans	sextant	10	0	Sextantis	Sex
Taurus ²	bull	5	+20	Tauri	Tau
Telescopium	telescope	19	−50	Telescopii	Tel
Triangulum	triangle	2	+30	Trianguli	Tri
Triangulum Australe	southern triangle	16	−65	Trianguli Australis	TrA
Tucana ¹⁷	toucan	0	−65	Tucanae	Tuc
Ursa Major	larger bear	11	+60	Ursae Majoris	UMa
Ursa Minor ¹⁸	smaller bear	16	+80	Ursae Minoris	UMi
Vela	sails	10	−45	Velorum	Vel
Virgo ²	virgin	13	0	Virginis	Vir
Volans	flying fish	8	−70	Volantis	Vol
Vulpecula	fox	20	+25	Vulpeculae	Vul

* Genitive is the grammatical case denoting possession. For example, astronomers denote the brightest or α (alpha) star in Orion (Betelgeuse) as α Orionis.

¹ Constellations of the area of the sky known as the wet quarter for its many watery images.

² A zodiac constellation.

³ Sometimes considered as Sagittarius's crown.

⁴ Ariadne's crown.

⁵ Corvus was Orpheus's companion, Lyra his harp.

⁶ Originally a part of Centaurus.

⁷ Contains the Large Magellanic Cloud and the south ecliptic pole.

⁸ Contains the north ecliptic pole.

⁹ One of the oldest constellations known.

¹⁰ Originally the claws of Scorpius.

¹¹ Originally named Musca Australis to distinguish it from Musca Borealis, the northern fly, which is now defunct; "Australis" is now dropped.

¹² Contains the south celestial pole.

¹³ Ophiucus is identified with the physician Aesculapius, and Serpens with the caduceus.

¹⁴ Contains the galactic center.

¹⁵ Originally named by Lacaille l'Atelier du Sculpteur (in Latin, Apparatus Sculptoris); now known simply as Sculptor. Contains the south galactic pole.

¹⁶ Shield of the Polish hero John Sobieski.

¹⁷ Contains the Small Magellanic Cloud.

¹⁸ Contains the north celestial pole.

APPENDIX I	SOME USEFUL ASTRONOMICAL QUANTITIES
Astronomical unit:	1 AU = 1.496×10^{11} m
Light-year:	1 ly = 9.461×10^{15} m = 63,240 AU
Parsec:	1 pc = 3.086×10^{16} m = 3.262 ly
Solar mass:	1 M _☉ = 1.989×10^{30} kg
Solar radius:	1 R _☉ = 6.960×10^8 m
Solar luminosity:	1 L _☉ = 3.827×10^{26} W
Earth's mass:	1 M _⊕ = 5.974×10^{24} kg
Earth's equatorial radius:	1 R _⊕ = 6.378×10^6 m
Moon's mass:	1 M _{Moon} = 7.349×10^{22} kg
Moon's equatorial radius:	1 R _{Moon} = 1.738×10^6 m

APPENDIX K	COMMON CONVERSIONS BETWEEN U.S. CUSTOMARY AND METRIC UNITS
	1 inch = 2.54 centimeters (cm)
	1 cm = 0.394 inch (in)
	1 yard = 0.914 meter (m)
	1 meter = 1.09 yards = 39.37 inches
	1 mile = 1.61 kilometers (km)
	1 km = 0.621 mile (mi)

APPENDIX J	SOME USEFUL PHYSICAL CONSTANTS
Speed of light:	$c = 2.998 \times 10^8$ m/s
Gravitational constant:	$G = 6.668 \times 10^{-11}$ N m ² kg ⁻²
Planck constant:	$h = 6.626 \times 10^{-34}$ J s = 4.136×10^{-15} eV s
Boltzmann constant:	$k = 1.380 \times 10^{-23}$ J K ⁻¹ = 8.617×10^{-5} eV K ⁻¹
Stefan–Boltzmann constant:	$\sigma = 5.670 \times 10^{-8}$ W m ⁻² K ⁻⁴
Mass of electron:	$m_e = 9.109 \times 10^{-31}$ kg
Mass of 1 H atom:	$m_H = 1.673 \times 10^{-27}$ kg

APPENDIX L MASS AND ENERGY INVENTORY FOR THE UNIVERSE

	Individual contribution	Section total
Dark matter and dark energy contributions		$0.954 \pm 0.003^*$
Dark energy	0.72 ± 0.03	
Dark matter	0.23 ± 0.03	
Primeval gravitational radiation	$\leq 10^{-10}$	
Contributions from Big Bang era		0.0010 ± 0.0005
Electromagnetic radiation	$10^{-4.3} \pm 0.000001$	
Neutrinos	$10^{-2.9} \pm 0.1$	
Normal particle (baryon) rest mass		0.045 ± 0.003
Charged particles (plasma) between stars and galaxies	0.0418 ± 0.003	
Main-sequence stars in elliptical galaxies and nuclear bulges	0.0015 ± 0.0004	
Neutral hydrogen and helium	0.00062 ± 0.00010	
Main-sequence stars in galactic disks and in irregular galaxies	0.00055 ± 0.00014	
White dwarfs	0.00036 ± 0.00008	
Molecular gas	0.00016 ± 0.00006	
Substellar objects	0.00014 ± 0.00007	
Black holes	0.00007 ± 0.00002	
Neutron stars	0.00005 ± 0.00002	
Planets	$10^{-6} \pm 0.1$	

*The numbers after the plus or minus ($\pm$) symbol indicate the possible errors in the given numbers.

This table lists the major contributions to the mass and energy of the universe.

APPENDIX M: READING GRAPHS

Reading Graphs

The graphs you will encounter in this book are compact ways of displaying patterns of information relating two variables, like the temperature and luminosity of stars or the temperature of an atmosphere at different altitudes. The relevant values of one of the variables are presented along the horizontal or x axis, and the relevant values of the other variable are presented along the vertical or y axis. The word “relevant” here indicates that often graphs do not start at 0. Consider three examples. First is data presented in Chapter 6 concerning the temperature of Venus’s atmosphere at different altitudes (Figure M-1).

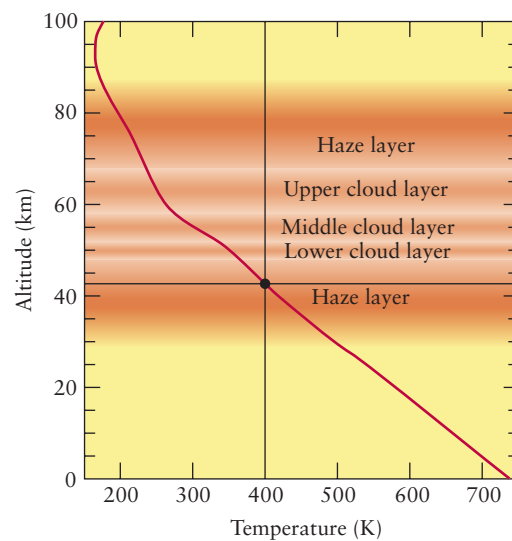


Figure M-1 Linear Graph

The horizontal axis of the graph indicates the temperature in kelvins (K). (Unlike degrees Celsius or degrees Fahrenheit, temperature using the Kelvin scale is simply noted as kelvins.)

The vertical axis denotes the altitude above Venus's surface in kilometers (km). Note that both axes are always labeled with a name (temperature or altitude, here) and units (K or km, respectively). Bear in mind that some variables in graphs you will see in this book increase to the right or upward (these are more common), but some variables will increase in the opposite directions. To read a graph, note that a value on the horizontal axis is transferred directly upward through the graph. For example, all points on the blue line in Figure M-1 (which extends upward from 400 K) have a temperature of 400 K. Equivalently, the value given on the vertical axis is transferred to all points horizontally across from this value. All the points on the black line in Figure M-1 are at an altitude of about 43 km above Venus's surface. We have interpolated between 40 and 45 to get this value (Figure M-2).

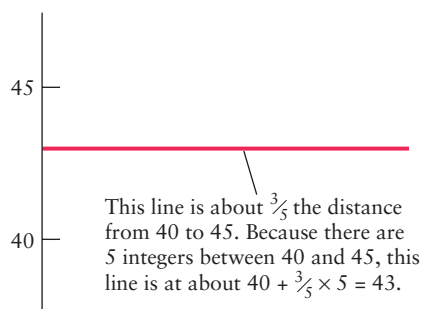


Figure M-2 Interpolation

A curve or a set of points on the graph presents the *relationship* between the variable represented on the horizontal axis and the variable on the vertical axis. Each point on a curve or each separate point relates the two variables. Choose a point, say, the dot on Figure M-1. It represents the temperature at a certain altitude above Venus's surface. To find the temperature at that point, you slide directly down (along the blue line in this example) from the point and read the value of the horizontal variable under it. To find the altitude for that point, you slide directly over to the side (along the black line) and read the value of the vertical variable there. In our example, sliding along the blue line leads to 400 K on the temperature line. Therefore, this point corresponds to a temperature of 400 K. Moving horizontally from the point, you encounter, by interpolation, the 43 km indicator. Combining the data, you conclude that *the temperature of Venus's atmosphere 43 km above its surface is 400 K.*

The red curve in Figure M-1 provides the relationship between altitude and temperature for a wide range of locations above Venus in a representation that is much more informative than a table of heights and temperatures. Specifically, this curve shows you the trend of temperature with altitude.

Try these questions: What is the temperature at 20 km? What is the altitude at which the temperature is 300 K? For most of this graph, what is the general *trend* of the temperature with height? (Answers appear at the end of the book.)

Sometimes, the known information is not a curve, but rather a set of points, as in our second example (Figure M-3) from Chapter 12. In this case, the graph connects the apparent magnitude of a nova (how bright it appears to be as seen from Earth regardless of its distance or other factors) and time. Each dot indicates how bright the nova (an explosion on the surface of certain stars) was at different times. For example, the peak brightness of the nova was an apparent magnitude of about +2 and it occurred on September 2. Noting that time passes to the right, you can immediately see that the trend of the nova's brightness is to increase rapidly and decrease more slowly.

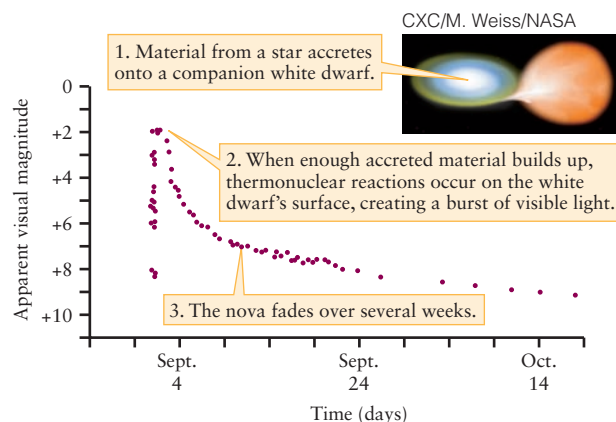


Figure M-3 The Brightness of a Nova

Try these questions: What is the apparent magnitude on September 24? October 9? On what two days was the apparent magnitude +6? (Answers appear at the end of the book.)

The graphs so far have shown variables that change uniformly along the axes. For example, the distance on Figure M-1 from 300 K to 400 K is the same as the distance from 400 K to 500 K, and so on. Many graphs you will encounter have variables that do not change uniformly (that is, linearly) along the axes. That means that the change in value going along each axis varies—there is not the same amount of change per centimeter

or inch along the axis. In **Figure M-4**, from Chapter 11, for example, the luminosity (total energy emitted per second) increases upward logarithmically, while the temperature decreases going from left to right in a more complex, nonuniform way.

The purpose of logarithmic and other nonlinear axes is to present in compact form data that vary very widely in values. For example, the dimmest star represented in Figure M-4 is just less than 0.1 times as luminous as the Sun, while the brightest star is nearly 1000 times as luminous. The process of getting information from logarithmic graphs is the same as linear graphs. You must just be careful not to think of values as doubling or tripling when you go over or up two or three

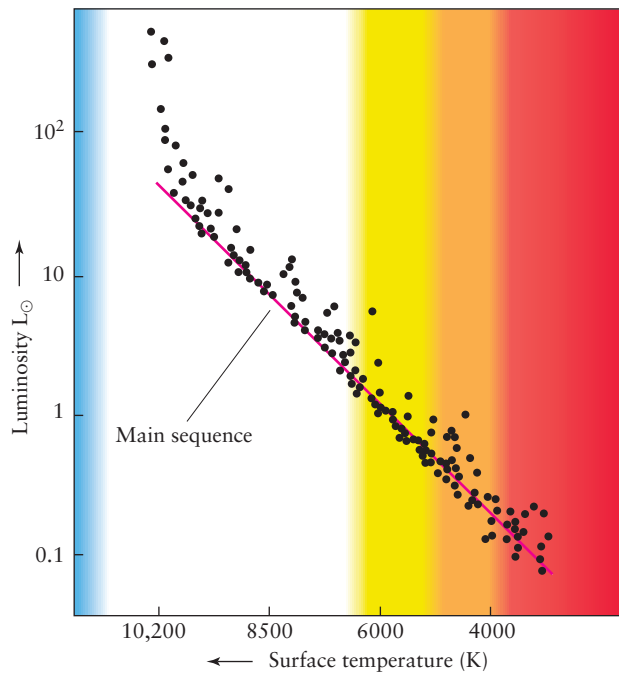


Figure M-4 Logarithmic Graph

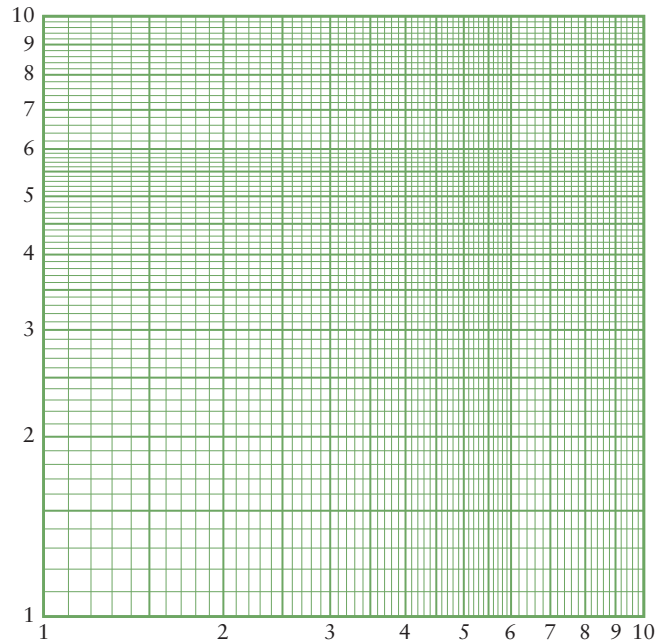


Figure M-5 Logarithmic Scale

intervals. **Figure M-5** shows how a logarithmic scale varies over one decade of values on both the vertical and horizontal axes. The same numbering intervals apply for any decade of values, for example, 1 to 10 or 10^5 to 10^6 . As you can see, the numbers bunch up near the highest value, so you need to interpolate these graphs more carefully than linear graphs.

Referring to Figure M-4, the luminosity of a star with surface temperature of 4000 K is about $0.1 L_{\odot}$. Note, also, that some graphs are linear on one axis and logarithmic on the other axis!

Try these questions: Approximately what is the luminosity of a star with surface temperature 8500 K? What is the surface temperature of a star with the same luminosity as the Sun ($1 L_{\odot}$)?

(Answers appear at the end of the book.)

APPENDIX N: PERIODIC TABLE OF THE ELEMENTS

1 H Hydrogen																	2 He Helium
3 Li Lithium	4 Be Beryllium											5 B Boron	6 C Carbon	7 N Nitrogen	8 O Oxygen	9 F Fluorine	10 Ne Neon
11 Na Sodium	12 Mg Magnesium											13 Al Aluminum	14 Si Silicon	15 P Phosphorus	16 S Sulfur	17 Cl Chlorine	18 Ar Argon
19 K Potassium	20 Ca Calcium	21 Sc Scandium	22 Ti Titanium	23 V Vanadium	24 Cr Chromium	25 Mn Manganese	26 Fe Iron	27 Co Cobalt	28 Ni Nickel	29 Cu Copper	30 Zn Zinc	31 Ga Gallium	32 Ge Germanium	33 As Arsenic	34 Se Selenium	35 Br Bromine	36 Kr Krypton
37 Rb Rubidium	38 Sr Strontium	39 Y Yttrium	40 Zr Zirconium	41 Nb Niobium	42 Mo Molybdenum	43 Tc Technetium	44 Ru Ruthenium	45 Rh Rhodium	46 Pd Palladium	47 Ag Silver	48 Cd Cadmium	49 In Indium	50 Sn Tin	51 Sb Antimony	52 Te Tellurium	53 I Iodine	54 Xe Xenon
55 Cs Cesium	56 Ba Barium	57 La Lanthanum	72 Hf Hafnium	73 Ta Tantalum	74 W Tungsten	75 Re Rhenium	76 Os Osmium	77 Ir Iridium	78 Pt Platinum	79 Au Gold	80 Hg Mercury	81 Tl Thallium	82 Pb Lead	83 Bi Bismuth	84 Po Polonium	85 At Astatine	86 Rn Radon
87 Fr Francium	88 Ra Radium	89 Ac Actinium	104 Rf Rutherfordium	105 Db Dubnium	106 Sg Seaborgium	107 Bh Bohrium	108 Hs Hassium	109 Mt Meitnerium	110 Ds Darmstadtium	111 Rg Roentgenium	112 Cn Copernicium	113 Uut Ununtrium	114 Fl Flerovium	115 Uup Ununpentium	116 Lv Livermorium	117 Uus Ununseptium	118 Uuo Ununoctium

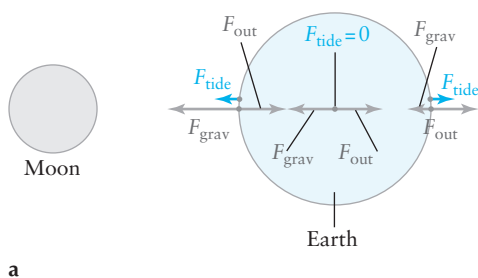
58 Ce Cerium	59 Pr Praseodymium	60 Nd Neodymium	61 Pm Promethium	62 Sm Samarium	63 Eu Europium	64 Gd Gadolinium	65 Tb Terbium	66 Dy Dysprosium	67 Ho Holmium	68 Er Erbium	69 Tm Thulium	70 Yb Ytterbium	71 Lu Lutetium
90 Th Thorium	91 Pa Protactinium	92 U Uranium	93 Np Neptunium	94 Pu Plutonium	95 Am Americium	96 Cm Curium	97 Bk Berkelium	98 Cf Californium	99 Es Einsteinium	100 Fm Fermium	101 Md Mendelevium	102 No Nobelium	103 Lr Lawrencium

APPENDIX O: TIDES

Tides result from a combination of Earth's motion around its barycenter with the Moon (or Sun) and the changing strength of the gravitational force from the Moon (or Sun) across Earth. Ignoring Earth's rotation and focusing on the Moon for the moment, let's first consider the motion of Earth and the Moon around their barycenter (see Figure 6-31a). The barycenter is located 1720 km (1068 mi) below Earth's surface on the line between the centers of Earth and the Moon. As a result of its motion around the barycenter, every point on Earth feels an equally strong force away from the Moon, unlike the outward force that you feel while riding on a merry-go-round, which increases with distance from the center. The force away from the Moon is represented at three locations by the force labeled F_{out} in Figure O-1a.

Opposing this force away from the Moon, Earth also has a gravitational attraction to its satellite. Recall from Section 2-8 on Newton's law of gravitation that the gravitational force exerted by one body on another decreases with distance. Therefore, the Moon's gravitational attraction on Earth is greatest at the point on Earth's surface closest to the Moon. This attraction is greater than the force felt at the center of Earth, which in turn is greater than the force felt on the opposite side of Earth from the Moon. The gravitational force in these places is labeled F_{grav} in Figure O-1a.

Because forces in the same or opposite directions simply add or subtract, respectively, we can subtract the two forces F_{out} and F_{grav} at any point on Earth, due to the presence of the Moon. The resulting net force is labeled F_{tide} in Figure O-1a and is the tidal force acting at these places. For example, at the point on Earth directly below the Moon, the tidal force is very strong and points toward the Moon. At the center of Earth, the two forces cancel each other completely. At the point on the opposite side of Earth from the Moon, the tidal force is equal in strength to the tidal force directly under the Moon but is directed away from the Moon.



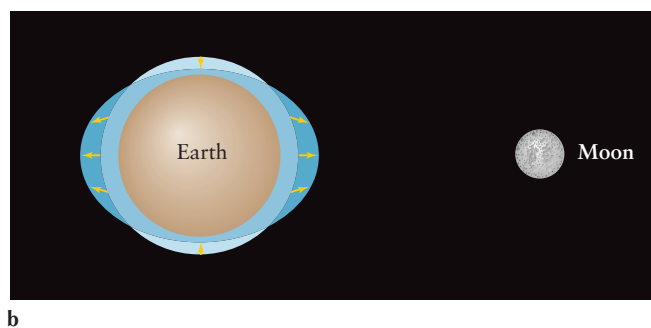
a

Figure O-1 Tidal Forces (a) The Moon induces tidal forces, F_{tide} , on Earth. At each point, this force is the difference between the force, F_{out} , created by the orbital motion of the two bodies around their barycenter, and the Moon's gravitational force, F_{grav} , at that point. The magnitude and direction of each arrow

Consider the point on Earth closest to the Moon. The strength of the tidal force, F_{tide} , seems to imply that the Moon's gravitational force lifts the water there to create the high tide. However, the force of gravitation from the Moon is not strong enough to raise that water more than a few centimeters or inches. Rather, the high tide closest to the Moon occurs because ocean waters from nearly halfway around to the opposite side of Earth are pulled Moonward by the gravitational force acting on them. This water slides toward the Moon and fills the ocean directly under it, creating a tidal bulge or high tide. Likewise, the net outward force acting on the opposite side of Earth from the Moon pushes the waters that are just over halfway around from the Moon to the side directly away from it, creating a simultaneous second high tide of equal magnitude on the opposite side of Earth from the Moon. Where the water has been drained away in this process, low tides occur.

Let's spin Earth back up. As it rotates, the geographic locations of the high and low tides continually change. Figure O-1b indicates that high tides occur when the Moon is high in the sky or below our feet, while low tides occur when the Moon is near either horizon. This timing means that there are two cycles of high and low tides at most places on Earth each day. These cycles do not span exactly 24 hours, however, because the Moon is moving along the celestial sphere, changing the time at which high or low tide occurs at any location from day to day. As a result, the time between consecutive high tides is 12 hours, 25 minutes.

Tides are complicated by the fact that the oceans are not uniformly deep, and the shores around continents and islands have a variety of shapes. As a result, in some places the oceans have only a single cycle of tides each day. Some places have tides that change negligibly, while some channels, such as the Bay of Fundy between the United States and Canada (maximum range between high and low tides of 16.3 m or 17.8 yd) and the Bristol Channel between Wales and England (maximum range of 15 m or 16.4 yd), have two tides that are each much higher than normal.



b

represent the strength and direction of each force. (b) Water slides along Earth to create the tides. Ignoring Earth's rotation and the effects of the continents for now, this figure shows how two high tides are created on Earth by the Moon's gravitational pull. The Sun has a weaker, but otherwise identical, effect.

Earth's rotation also has the effect of moving the high tides away from the line between the centers of Earth and the Moon. Earth's eastward rotation is nearly 30 times faster than the Moon's eastward revolution around Earth. Therefore, as it rotates, Earth drags the high tides eastward with it. As a result, the high tide that should be directly between the centers of Earth and the Moon is actually 10° ahead of the Moon in its orbit around Earth (Figure O-2). This high tide stays at roughly this position because the Moon's gravitational force pulls the tide westward while it is being dragged eastward by Earth's rotation; the two effects balance each other out. A similar argument applies to the tide on the opposite side of Earth.

We can repeat this analysis for the Earth–Sun system and get the same qualitative results.

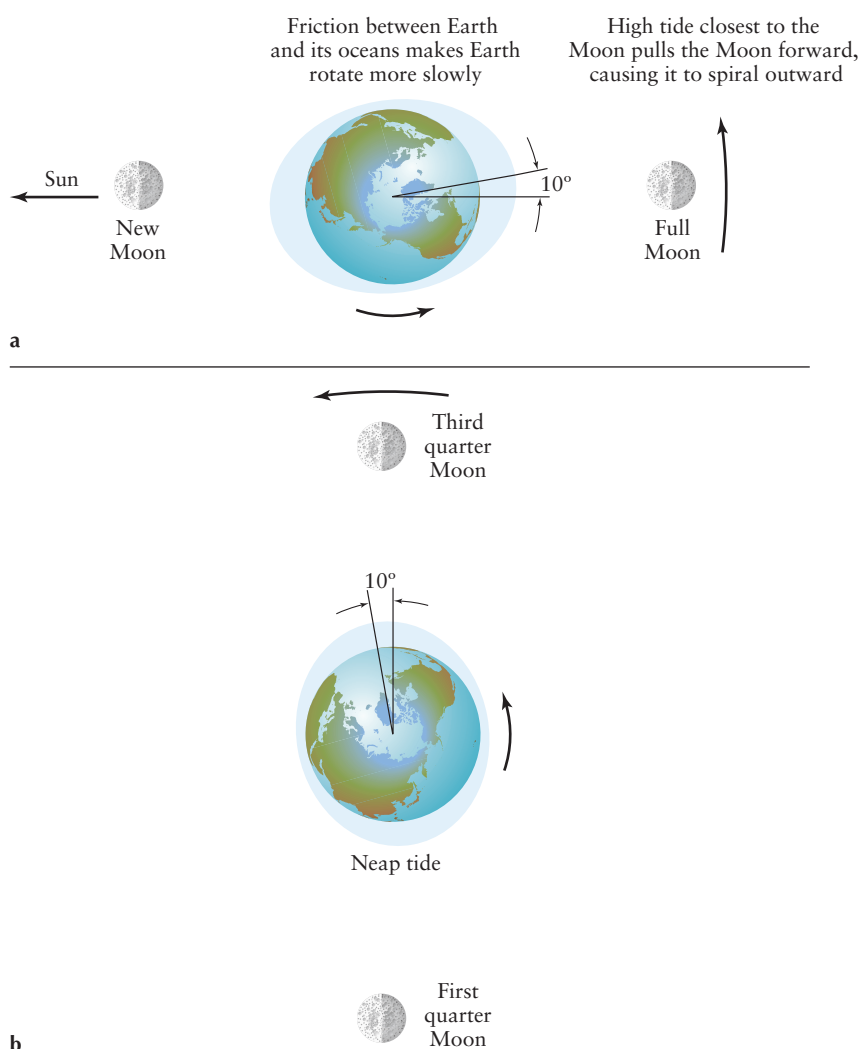


Figure O-2 Tides on Earth The gravitational forces of the Moon and the Sun deform the oceans. Due to Earth's rapid rotation, the high tide closest to the Moon leads it around Earth by about 10° . **(a)** The greatest deformation (spring tides) occurs when the Sun, Earth, and the Moon are aligned with the Sun and the Moon either on the same or opposite sides of Earth. **(b)** The least deformation (neap tides) occurs when the Sun, Earth, and the Moon form a right angle.

APPENDIX P: ENERGY AND MOMENTUM

Scientists identify two types of energy that are available to any object. The first, called **kinetic energy**, is associated with the object's motion. For speeds much less than the speed of light, we can write the amount of kinetic energy, KE, of an object as

$$KE = \frac{1}{2} mv^2$$

where m is the object's mass and v is its speed. Kinetic energy is a measure of how much work the object can do on the outside world or, equivalently, how much work the outside world has done to give the object this speed. Kinetic energy applies to rotation and revolution, as well as to straight-line motion.

Work is also a rigorously defined concept that often is at odds with our intuition. Work is defined as the product of the force, F , acting on an object and the distance, d , over which the object moves in the direction of the force:

$$W = Fd$$

For example, if I exert a horizontal force of 50 N (N is the unit newtons and is the metric unit of force) and thereby move an object 10 m in the direction I push it, then I have done $50 \text{ N} \times 10 \text{ m} = 500 \text{ J}$ of work. (I have used the relationship that 1 newton $\times$ 1 meter = 1 joule.)

The second type of energy is called **potential energy**, the energy available to an object as a result of its location in space. For example, if you hold a pencil above the ground, the pencil has potential energy that can be converted into kinetic energy by Earth's gravitational force. How does that conversion get underway? Just let go of the pencil.

There are various kinds of potential energy, such as the potential energy stored in a battery and the potential energy stored in objects under the influence of gravity. We will focus on *gravitational potential energy*. Far from extremely massive objects, like stars, or extremely dense objects, like black holes, gravitational potential energy, PE, can be written as

$$PE = \frac{GmM}{r}$$

where the constant $G = 6.6683 \times 10^{-11} \text{ N m}^2/\text{kg}^2$, m is the mass of the object whose gravitational potential energy you are measuring, M is the mass of the object generating the gravitational attraction, and r is the distance between the centers of mass of these two objects.

Near the surface of Earth, this equation simplifies to

$$PE = mgh$$

where $g = 9.8 \text{ m/s}^2$ (32 ft/s^2) is the gravitational acceleration at Earth's surface, and h is the height of the object above Earth's surface.

Potential energy can be converted into kinetic energy and vice versa. After you drop a pencil, its gravitational potential energy begins to decrease while its kinetic energy begins to increase at the *same rate*. The pencil's total energy is conserved. Conversely, if you throw a pencil up in the air, the kinetic energy you give it will immediately begin to decrease, while its potential energy increases at the same rate.

Related to the motion of an object and, hence, to its kinetic energy are the concepts of *linear momentum*, usually just called **momentum**, and **angular momentum**. Momentum, p , is described by the equation

$$p = mv$$

where v is the velocity of the object. Both p and v are in boldface to indicate that they both represent motion in some direction or another (both in the same direction), as well as a numeric value. Simple algebra reveals that kinetic energy and momentum are related by

$$\text{KE} = \frac{p^2}{2m}$$

Linear momentum, then, indicates how much energy is available to an object because of its motion in a straight line (linear motion).

Angular momentum, L , can be expressed mathematically as

$$L = I\omega$$

where I is the **moment of inertia** of an object, and ω (lowercase Greek omega) is the angular speed and direction of the rotating object. Just as an object's mass indicates how hard it is to change an object's straight-line motion, the moment of inertia indicates how hard it is to change the rate at which an object rotates or revolves. The moment of inertia depends on an object's mass and shape. Kinetic energy due to angular motion can be written as

$$\text{KE} = \frac{L^2}{2I}$$

Newton's first law can also be expressed in terms of **conservation of linear momentum**:

A body maintains its linear momentum unless acted upon by a net external force.

Equivalently, for angular motion we can write the **conservation of angular momentum**:

A body maintains its angular momentum unless acted upon by a net external torque.

Torques are created when a force acts on an object in some direction other than toward the center of the object's angular motion, as shown in **Figure P-1**.

Earth has angular momentum from two sources, namely, from spinning on its rotation axis and from orbiting the Sun. Likewise, the Moon has angular momentum because it spins on its rotation axis and it orbits Earth. Virtually all objects in astronomy have angular momentum, and it is probably fair to say that conservation of angular momentum is among the most important laws in the cosmos. After all, conservation of angular momentum is what keeps the planets in orbit around the Sun, the moons in orbit around the planets, and astronomical bodies rotating at relatively constant rates, as well as causing many other rotation-related effects that we will encounter throughout this book.

Try these questions: How does tripling the linear momentum of an object change its kinetic energy? How does halving the angular momentum of an object change its kinetic energy? How much work would you do if you pushed on a desk with a force of 100 N, while it moved 20 m? How much work would you do if you pushed on a desk with a force of 500 N, and it moved 0 m? What two things can you vary to change the angular momentum of an object?

(Answers appear at the end of the book.)

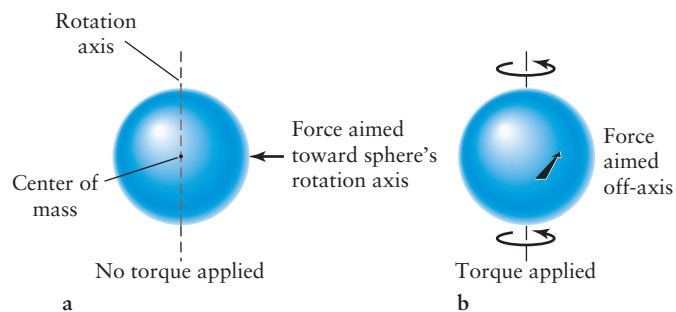


Figure P-1 Angular Momentum and Torque (a) When a force acts through an object's rotation axis or toward its center of mass, the force does not exert a torque on the object. (b) When a force acts in some other direction, then it exerts a torque, causing the body's angular momentum to change. If the object can spin around a fixed axis, like a globe, then the rotation axis is the rod running through it. If the object is not held in place, then the rotation axis is in a line through a point called the object's *center of mass*. The center of mass of any object is the point that follows a smooth, elliptical path as the object moves in response to a gravitational field. All other points in the spinning object wobble as it moves.

APPENDIX Q: RADIOACTIVITY AND THE AGES OF OBJECTS

The isotopes of many elements are radioactive, meaning that the elements spontaneously transform into other elements. Each radioactive isotope has a distinctive half-life, which is the time that it takes half of the initial concentration of the isotope to transform into another element. After two half-lives, a radioactive isotope is reduced to $\frac{1}{2} \times \frac{1}{2}$ or $\frac{1}{4}$ of its initial concentration (see **Figure Q-1**). Among the most important radioactive elements for determining the age of objects in astronomy is the isotope of uranium with 146 neutrons, ^{238}U (“U two-thirty-eight”). The half-life of ^{238}U decaying into lead is 4.5 billion years.

To determine the time since an object, such as a piece of space debris discovered on Earth, solidified, scientists estimate how much lead the object had when it formed. Then they measure the amounts of uranium and lead that it contains now. Subtracting the amount of original lead, they use the amount of uranium and lead, together with the graph of radioactive decay, to determine the object’s age.

Example: Suppose a piece of space debris discovered on Earth was determined to have equal amounts of lead and uranium. How long ago did this debris form? Assuming that it originally had no lead, we see from the chart that a 1-to-1 mix of lead to uranium occurs 4.5 billion years after the object formed. This period is one half-life of uranium.

Compare! This process works with any radioactive isotope. However, some isotopes have such short half-lives that they are not useful in astronomy. For example, the well-known carbon dating used to determine the

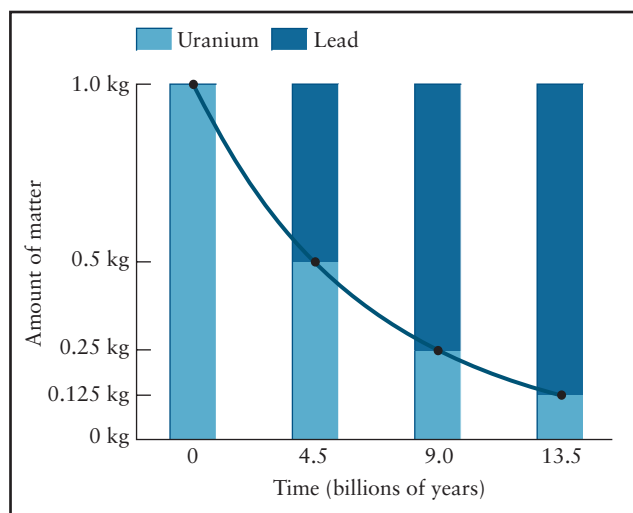


Figure Q-1 The Transformation of Uranium into Lead This figure shows the rate that 1 kg of uranium decays into lead, as described in the text. The 1-kg sample contains 0.125 kg of uranium after 13.5 billion years.

ages of ancient artifacts on Earth is of little use in astronomy because ^{14}C has a half-life of only 5730 years. Because so much of the carbon in a sample has decayed away by then, carbon dating is useful only for time intervals shorter than 100,000 years, usually a period over which little of astronomical importance occurs.

Try these questions: What fraction of a kilogram of radioactive material remains after three half-lives have passed? Using Figure Q-1, estimate how much uranium will remain after 6 billion years. Approximately how many half-lives of ^{14}C pass in 100,000 years? (Answers appear at the end of the book.)

APPENDIX R: GRAVITATIONAL FORCE

From Newton’s law of gravitation, if two objects that have masses m_1 and m_2 are separated by a distance r , then the gravitational force, F , between them is

$$F = \frac{Gm_1m_2}{r^2}$$

In this formula, G is the **universal constant of gravitation**.

The equation $F = G(m_1m_2/r^2)$ gives, for example, the force from the Sun on Earth and, equivalently, the force from Earth on the Sun. If m_1 is the mass of Earth (6.0×10^{24} kg), m_2 is the mass of the Sun (2.0×10^{30} kg), and r is the distance from the center of Earth to the center of the Sun (1.5×10^{11} m):

$$F = 3.6 \times 10^{22} \text{ N}$$

where N is the unit of force, a newton. This number can then be used in Newton’s second law, $F = ma$, to find the acceleration of Earth due to the Sun. This yields

$$a_{\text{Earth}} = F/m_{\text{Earth}} = 6.0 \times 10^{-2} \text{ m/s}^2$$

Newton’s third law says that Earth exerts the same force on the Sun, so the Sun’s acceleration due to Earth’s gravitational force is

$$a_{\text{Sun}} = F/m_{\text{Sun}} = 1.8 \times 10^{-8} \text{ m/s}^2$$

In other words, Earth pulls on the Sun, causing the Sun to move toward it. Because of the Sun’s greater mass, however, the amount that the Sun accelerates Earth is more than 300,000 times greater than the amount that Earth accelerates the Sun.

Try these questions: Earth’s radius is 6.4×10^6 m and 1 kg is a mass equivalent to a weight of 9.8 N (or 2.2 lb) on Earth. What is the force that Earth exerts on you in newtons and pounds? What is the force that you exert on Earth in these units? What would the Sun’s force (in newtons) be on Earth if our planet were twice as far from the Sun as it is? How does that force compare to the force from the Sun at our present location? (Answers appear at the end of the book.)

APPENDIX S LARGEST OPTICAL TELESCOPES IN THE WORLD

Aperture (m)	Name	Location	Altitude (m)
10.4	Gran Telescopio Canarias	La Palma, Canary Islands, Spain	2400
10	Keck I	Mauna Kea, Hawaii	4123
10	Keck II	Mauna Kea, Hawaii	4123
10	SALT	Sutherland, South Africa	1759
9.2	Hobby–Eberly	Mt. Fowlkes, Texas	2072
8.4 × 2	Large Binocular Telescope	Mt. Graham, Arizona	3170
8.3	Subaru	Mauna Kea, Hawaii	4100
8.2	Antu	Cerro Paranal, Chile	2635
8.2	Kueyen	Cerro Paranal, Chile	2635
8.2	Melipal	Cerro Paranal, Chile	2635
8.2	Tepun	Cerro Paranal, Chile	2635
8.1	Gemini North (Gillett)	Mauna Kea, Hawaii	4100
8.1	Gemini South	Cerro Pachon, Chile	2737
6.5	MMT	Mt. Hopkins, Arizona	2600
6.5	Walter Baade	La Serena, Chile	2282
6.5	Landon Clay	La Serena, Chile	2282
6	Bolshoi Teleskop Azimutalnyi	Nizhny Arkhyz, Russia	2070
6	LZT	British Columbia, Canada	395
5	Hale	Palomar Mtn., California	1900
4.2	William Herschel	La Palma, Canary Islands, Spain	2400
4.2	SOAR	Cerro Pachon, Chile	2738
4.2	LAMOST	Xinglong Station, China	950
4	Victor Blanco	Cerro Tololo, Chile	2200
3.9	Anglo–Australian	Coonabarabran, Australia	1100
3.8	Mayall	Kitt Peak, Arizona	2100
3.8	UKIRT	Mauna Kea, Hawaii	4200
3.7	AEOS	Maui, Hawaii	3058
3.6	“360”	Cerro La Silla, Chile	2400
3.6	Canada–France–Hawaii	Mauna Kea, Hawaii	4200
3.6	Telescopio Nazionale Galileo	La Palma, Canary Islands, Spain	2387
3.5	MPI–CAHA	Calar Alto, Spain	2200
3.5	New Technology	Cerro La Silla, Chile	2400
3.5	ARC	Apache Point, New Mexico	2788
3.5	WIYN	Kitt Peak, Arizona	2100
3.5	Starfire	Kirtland Air Force Base, New Mexico	1900
3	Shane	Mt. Hamilton, California	1300
3	NASA IRTF	Mauna Kea, Hawaii	4160
2.7	Harlan Smith	Mt. Locke, Texas	2100
2.6	BAO	Byurakan, Armenia	1405

APPENDIX S

LARGEST OPTICAL TELESCOPES IN THE WORLD (Continued)

Aperture (m)	Name	Location	Altitude (m)
2.6	Shajn	Crimea, Ukraine	600
2.5	Hooker	Mt. Wilson, California	1700
2.5	Isaac Newton	La Palma, Canary Islands, Spain	2382
2.5	Nordic Optical	La Palma, Canary Islands, Spain	2382
2.5	duPont	La Serena, Chile	2282
2.5	Sloan Digital Sky Survey	Apache Point, New Mexico	2788
2.4	Hiltner	Kitt Peak, Arizona	2100
2.4	Lijiang	Lijiang City, China	3250
2.4	Hubble Space Telescope	Low Earth orbit	6×10^5
2.3	WIRO	Jelm Mtn., Wyoming	2900
2.3	ANU	Coonabarabran, Australia	1100
2.3	Bok	Kitt Peak, Arizona	2100
2.3	Vainu Bappu	Kavalur, India	700
2.3	Aristarchos	Mt. Helmos, Greece	2340
2.2	ESO–MPI	Cerro La Silla, Chile	2335
2.2	MPI–CAHA	Calar Alto, Spain	2200
2.2	UH	Mauna Kea, Hawaii	4200
2.1	Kitt Peak 2.1 meter	Kitt Peak, Arizona	2100
2.1	Otto Struve	Davis Mountains, Texas	2070
2.1	UNAM	San Pedro, Mexico	2800
2.1	Jorge Sahade	El Leoncito, Argentina	2552
2	Himalayan Chandra	Hanle, India	4517
2	Alfred Jensch Teleskop	Tautenburg, Germany	
2	Carl Zeiss Jena	Azerbaijan	
2	Ondrejov	Ondrejov, Czech Republic	
2	RCC	Chepelare, Bulgaria	
2	Bernard Lyot	Pic du Midi, France	2877
2	Faulkes Telescope North	Haleakala, Maui, Hawaii	3050
2	Faulkes Telescope South	Siding Springs, Australia	3822
2	MAGNUM	Haleakala, Maui, Hawaii	3058
1.0 (× 6)	CHARA Array	Mt. Wilson, California	1700

GLOSSARY

A ring The outermost ring of Saturn visible from Earth; it is located just beyond the Cassini division.

absolute magnitude The apparent magnitude that a star would have if it were 10 parsecs from Earth.

absorption line A dark line in a continuous spectrum created when photons of a certain energy are absorbed by atoms or molecules.

absorption line spectrum Dark lines superimposed on a continuous spectrum.

acceleration A change in the direction or magnitude of a velocity.

accretion The gradual accumulation of matter by an astronomical body, usually caused by gravity.

accretion disk An orbiting disk of matter spiraling in toward a star or black hole.

active galactic nuclei (AGN) Supermassive black holes in the cores of some galaxies that emit particles and radiation that, when viewed from different angles, create Seyfert galaxies, radio galaxies, double radio sources, BL Lacertae objects, and quasars.

active galaxy A very luminous galaxy, often containing an active galactic nucleus.

active optics A system that adjusts a reflecting telescope in response to changes in temperature and shape of the mount; it helps optimize an image.

adaptive optics Primary telescope mirrors that are continuously and automatically adjusted to compensate for the distortion of starlight due to the motion of Earth's atmosphere.

AGB star See **asymptotic giant branch (AGB) star**.

albedo The fraction of sunlight that a planet, asteroid, or satellite scatters directly back into space.

altitude The angle straight upward from the horizon to the object under observation.

amino acids A class of chemical compounds that are the building blocks of proteins.

angle The opening between two straight lines that meet at a point.

angular diameter (angular size) The arc angle across an object.

angular momentum A measure of how much energy an object has stored in its rotation and/or revolution.

angular resolution The angular size of the smallest detail of an astronomical object that can be distinguished with a telescope.

annular eclipse An eclipse of the Sun in which the Moon is too distant to cover the Sun completely so that a ring of sunlight is seen around the Moon at mid-eclipse.

anorthosite A light-colored rock found throughout the lunar highlands and in some very old mountains on Earth.

aphelion The point in its orbit where a planet or other solar system body is farthest from the Sun.

Apollo asteroid An asteroid that is sometimes closer to the Sun than Earth is.

apparent magnitude A measure of the brightness of light from a star or other object as seen from Earth.

arc angle The measurement of the angle between two objects or two parts of the same object.

asterism Easily recognizable pattern of stars.

asteroid Any of the rocky objects larger than about ten meters (or ten yards) in diameter (and not classified as a planet or moon) that orbits the Sun.

asteroid belt A 1½-AU-wide region between the orbits of Mars and Jupiter in which most of the asteroids are found.

astrobiology The study of life in the universe.

astronomical unit (AU) The average distance between Earth and the Sun: 1.5×10^8 km = 93 million mi.

asymptotic giant branch (AGB) star A red giant star that has completed core helium fusion and has reexpanded for a second time.

atomic number The number of protons in the nucleus of an atom.

autumnal equinox The intersection of the ecliptic and the celestial equator where the Sun crosses the equator moving from north to south. The beginning of autumn (around September 23).

average density The mass of an object divided by its volume.

azimuth Angle eastward from due north along a viewer's horizon.

B ring The brightest of the three rings of Saturn visible from Earth; it lies just inside the Cassini division.

barred spiral galaxy A spiral galaxy in which the spiral arms begin from the ends of a bar running through the nuclear bulge.

belt asteroid An asteroid whose orbit lies in the asteroid belt.

belts Relatively dark bands running east-west in the cloud cover of the four giant planets.

Big Bang An explosion that took place roughly 13.8 billion years ago, creating all space, time, matter, and energy in which the universe emerged.

binary star Two stars revolving about each other; a double star.

birth line A line on the Hertzsprung-Russell diagram corresponding to where stars with different masses transform from protostars to pre-main-sequence stars.

BL Lacertae (BL Lac) object A type of active galaxy; a blazar.

black hole An object whose gravity is so strong that the escape velocity from inside it exceeds the speed of light.

blackbody A hypothetical perfect radiator that absorbs and reemits all radiation falling upon it.

blackbody curve The curve obtained when the intensity of radiation from a blackbody at a particular temperature is plotted against wavelength.

blazar A BL Lacertae object.

blueshift A shift of all spectral features toward shorter wavelengths; the Doppler shift of light from an approaching source.

Bok globule A small, roundish, dark nebula in which stars are forming.

brown dwarf Any of the planetlike bodies with less than $0.08 M_{\odot}$ and more than about $13 M_{\text{Jupiter}}$; such bodies do not have enough mass to sustain fusion in their cores.

C ring The faint, inner portion of Saturn's main ring system.

caldera The crater at the summit of a volcano.

carbonaceous chondrites A class of extremely ancient, carbon-rich meteorites.

Cassegrain focus An optical arrangement in a reflecting telescope in which light rays are reflected by a secondary mirror through a hole in the primary mirror.

Cassini division A prominent gap between Saturn's A and B rings discovered in 1675 by Giovanni Cassini.

celestial equator A great circle on the celestial sphere 90° from the celestial poles.

celestial sphere A hypothetical sphere of very large radius centered on the observer; the apparent sphere of the night sky.

Celsius temperature scale (also called centigrade) Temperature scale with 0°C at the freezing point of water and 100°C at the boiling point of water.

center of mass The point around which a rigid system is perfectly balanced in a gravitational field; also, the point in space around which mutually orbiting bodies have elliptical orbits.

central bulge A flattened sphere of stars centered on a spiral galaxy's nucleus, extending out to the vicinity of the spiral arms.

Cepheid variable One of two types of yellow, supergiant, pulsating stars.

Cerenkov radiation Radiation produced by particles traveling through a substance faster than light can.

Chandrasekhar limit The maximum mass of a white dwarf, about $1.4 M_{\odot}$.

charge-coupled device (CCD) A type of solid-state silicon wafer designed to detect photons.

chondrites Stony meteorites that have never been melted or otherwise altered since they formed. They contain a variety of debris, including spherical droplets called chondrules.

chromatic aberration An optical property whereby different colors of light passing through a lens are focused at different distances from it.

chromosphere The layer in the solar atmosphere between the photosphere and the corona.

circumpolar stars All the stars that never set at a given latitude; all the stars between Polaris and the northern horizon.

close binary A binary star system whose members are separated by a few stellar diameters.

closed universe A universe that contains enough matter to cause it to recollapse. It is finite in extent and has no "outside."

cluster (of galaxies) A collection of a few hundred to a few thousand galaxies bound by gravity.

collision-ejection theory The theory that the Moon was created by the impact of a planet-sized object with Earth; presently considered the most plausible theory of the Moon's formation.

coma (of a comet) The nearly spherical, diffuse gas surrounding the nucleus of a comet near the Sun.

comet A small body of ice and dust in orbit about the Sun. While passing near the Sun, a comet's vaporized ices give rise to a coma, tails, and a hydrogen envelope.

configuration (of a planet) A particular geometric arrangement of Earth, a planet, and the Sun.

confinement The moment shortly after the Big Bang when quarks bound together to form particles like protons and neutrons.

conjunction The alignment of two bodies in the solar system so that they appear in the same part of the sky as seen from Earth.

conservation of angular momentum The law of physics stating that the total amount of angular momentum in an isolated system remains constant.

conservation of linear momentum If the sum of the external forces on a system remains zero, the total linear momentum of the system remains constant.

constellation Any of the 88 contiguous regions that cover the entire celestial sphere, including all the objects in each region; also, informally, a configuration of stars often named after an object, a person, or an animal (more properly called an asterism).

contact binary A close binary system in which both stars fill or overflow their Roche lobes.

continental drift The gradual movement of the continents over the surface of Earth due to plate tectonics.

continuous spectrum (continuum) A spectrum of light over a range of wavelengths without any spectral lines.

convection The transfer of energy by moving currents of fluid or gas containing that energy.

convective zone A layer in a star where energy is transported outward by means of convection; also known as the *convective envelope* or *convection zone*.

core The central portion of any astronomical object.

core (of the Sun) The central quarter of the Sun's radius, in which hydrogen is fused into helium, releasing the energy that enables the Sun to shine.

core helium fusion The fusion of helium to form carbon and oxygen at the center of a star.

corona The Sun's outer atmosphere.

coronal hole A relatively dark region of the Sun's inner corona as seen at X-ray wavelengths.

coronal mass ejection Large volumes of high-energy gas released from the Sun's corona.

cosmic censorship The belief that the only connection between a black hole and the universe is the black hole's event horizon.

cosmic light horizon A sphere, centered on Earth, whose radius equals the distance traveled by light since the Big Bang.

cosmic microwave background Photons from every part of the sky with a blackbody spectrum at 2.73 K; the cooled-off radiation from the primordial fireball that originally filled all space.

cosmic rays (primary cosmic rays) High-speed particles traveling through space.

cosmic ray shower Groups of particles from Earth's atmosphere propelled Earthward by the impact of a cosmic ray; *see also the related secondary cosmic rays*.

cosmological constant A number sometimes inserted into the equations of general relativity that represents a pressure that opposes gravity throughout the universe.

cosmological redshift An increase in wavelength from distant galaxies and quasars caused by the expansion of the universe.

cosmology The study of the evolution and large-scale structure of the universe.

coudé focus A reflecting telescope in which a series of mirrors directs light to a remote focus away from the moving parts of the telescope.

crater A circular depression on a celestial body caused by the impact of a meteoroid, asteroid, or comet or by a volcano.

crust The solid surface layer of some astronomical bodies, including the terrestrial planets, the moons, the asteroids, and some stellar remnants.

dark ages The age of the universe between the time of decoupling and the first burst of star formation.

dark energy A repulsive gravitational effect that is causing the universe to accelerate outward.

dark matter The as-yet-undetected matter in the universe that is underluminous and probably quite different from ordinary matter.

dark nebula A cloud of interstellar gas and dust that obscures the light of more distant stars.

declination (Dec.) The coordinate on the celestial sphere exactly analogous to latitude on Earth; measured north and south of the celestial equator.

decoupling The epoch in the early universe when electrons and ions first combined to create stable atoms; the time when electromagnetic radiation ceased to dominate over matter.

degree (°) A unit of angular measure or a temperature measure.

dense core Any of the regions of interstellar gas clouds that are slightly denser than normal and destined to collapse to form one or a few stars.

detached binary A binary system in which the surfaces of both stars are inside their respective Roche lobes.

differential rotation The rotation of a nonrigid object in which parts at different latitudes or different radial distances move at different speeds.

diffraction grating An optical device consisting of closely spaced lines ruled on a piece of glass that is used like a prism to disperse light into a spectrum.

direct motion The gradual, eastward apparent motion of a planet against the background stars as seen from Earth.

disk (of a galaxy) A flattened assemblage of stars, gas, and dust in a spiral galaxy like the Milky Way, or in a lenticular galaxy.

diurnal motion Cyclic motion with a 1-day period.

Doppler shift The change in wavelength of radiation due to relative motion between and the source and the observer directly toward or away from each other.

double radio source An extragalactic radio source characterized by two large lobes of radio emission, often located on either side of an active galaxy.

Drake equation A mathematical equation used to estimate the number of extraterrestrial civilizations that may exist in our Galaxy.

dust devil Whirlwind found in dry or desert areas on both Earth and Mars.

dust tail A comet tail composed of dust particles escaping from the comet's nucleus.

dwarf planet A celestial body that is in orbit around the Sun and has sufficient mass for its self-gravity to pull itself into a nearly spherical shape but does not have enough gravity to clear its orbital neighborhood of all the small debris orbiting there.

dynamo theory The generation of a magnetic field by circulating electric charges.

eclipse path The track of the tip of the Moon's shadow along Earth's surface during a total or annular solar eclipse.

eclipsing binary A double star system in which stars periodically pass in front of each other as seen from Earth.

ecliptic The annual path of the Sun on the celestial sphere; the plane of Earth's orbit around the Sun.

Einstein cross The appearance of four images of the same galaxy or quasar due to gravitational lensing by an intervening galaxy.

Einstein ring The circular or arc-shaped image of a distant galaxy or quasar created by gravitational lensing by an intervening galaxy.

ejecta blanket The ring of material surrounding a crater that was ejected during the crater-forming impact.

electromagnetic force The interaction between charged particles, one of the four fundamental forces in nature.

electromagnetic radiation Radiation consisting of oscillating electric and magnetic fields, namely gamma rays, X-rays, visible light, ultraviolet and infrared radiation, microwaves, and radio waves.

electromagnetic spectrum The entire array of electromagnetic radiation.

electron degeneracy pressure A powerful pressure produced by repulsion of closely packed (degenerate) electrons.

element A substance that cannot be decomposed by chemical means into simpler substances. Every atom of the same element contains the same number of protons.

elliptical galaxy A galaxy with an elliptical shape, little interstellar matter, and no spiral arms.

elongation The angle between a planet and the Sun as seen from Earth.

emission line A bright line of electromagnetic radiation.

emission line spectrum A spectrum that contains only bright emission lines.

emission nebula A glowing gaseous nebula whose light comes from fluorescence caused by a nearby star.

Encke division A thin gap in Saturn's A ring, possibly first seen by Johann Franz Encke in 1838.

equinox Either of the two days of the year when the Sun crosses the celestial equator and is therefore directly over Earth's equator; *see also* **autumnal equinox** and **vernal equinox**.

era of recombination The time, roughly 380,000 years after the Big Bang, when the universe became transparent.

ergoregion The region of space immediately outside the event horizon of a rotating black hole where it is impossible to remain at rest.

event horizon The location around a black hole where the escape velocity equals the speed of light; the boundary of a black hole.

evolutionary track On the Hertzsprung-Russell diagram, the path followed by a point representing an evolving star.

exoplanet A planet orbiting a star other than the Sun.

expanding universe The motion of the superclusters of galaxies away from each other.

eyepiece lens A magnifying lens used to view the image produced at the focus of a telescope.

F ring A thin ring just beyond the outer edge of Saturn's main ring system.

filament A dark curve seen above the Sun's photosphere that is the top view of a solar prominence.

flare *See* **solar flare**.

focal length The distance from a lens or concave mirror to where converging light rays meet.

focal plane The plane at the focal length of a lens or concave mirror on which an extended object is focused.

focal point The place at the focal length where light rays from a point object (that is, one that is too distant or tiny to resolve) are converged by a lens or concave mirror.

focus (*plural foci*) (**of an ellipse**) The two points inside an ellipse the sum of whose distances from any point on the ellipse is constant.

force That which can change the momentum of an object.

frequency The number of peaks or troughs of a wave that pass a fixed point each second. Also, the number of complete vibrations or oscillations per second.

galactic cannibalism A collision between two galaxies of unequal mass and size in which the smaller galaxy is absorbed by the larger galaxy.

galactic merger A collision and subsequent merger of two roughly equal-sized galaxies.

galactic nucleus The center of a galaxy; the center of the Milky Way Galaxy.

galaxy A large assemblage of stars, gas, and dust bound together by their mutual gravitational attraction.

Galilean moon (Galilean satellite) Any one of the four large moons of Jupiter (Io, Europa, Ganymede, Callisto) that is visible from Earth through a small telescope.

gamma ray The most energetic form of electromagnetic radiation.

gamma-ray burst A short burst of gamma rays; the sources of the bursts are outside our Galaxy.

gas (ion) tail The relatively straight tail of a comet produced by the solar wind acting on ions in a comet's coma.

giant molecular cloud A large interstellar cloud of cool gas and dust in a galaxy.

giant star A star whose diameter is roughly 10 to 100 times that of the diameter of the Sun.

glitch A sudden speedup in the period of a pulsar.

globular cluster A large spherical cluster of gravitationally bound stars usually found in the outlying regions of a galaxy.

Grand Unified Theories (GUTs) Theories that each describe and explain electromagnetism, the weak forces, and the strong forces.

granules Lightly colored convection features about 1000 km (620 mi) in diameter seen constantly in the solar photosphere.

gravitation The tendency of all matter to attract all other matter.

gravitational lensing The distortion of the appearance of an object by a source of gravity between it and the observer.

gravitational radiation Energy that travels at the speed of light as a result of the geometry of space changing shape.

gravitational redshift The redshift of photons leaving the gravitational field of any massive object, such as a star or black hole.

gravitational waves Ripples in the overall geometry of space, traveling at the speed of light, that are produced as any object changes shape.

gravity *See* gravitation.

Great Dark Spot A large, dark, oval-shaped storm that used to be in Neptune's southern hemisphere.

Great Red Spot A large, red-orange, oval-shaped storm in Jupiter's southern hemisphere.

greenhouse effect The trapping of infrared radiation near a planet's surface by the planet's atmosphere.

H II region A region of ionized hydrogen in interstellar space.

habitable zone The region around any star wherein water can exist in liquid form on a planet's surface and, hence, life as we know it can conceivably exist.

halo (of a galaxy) A spherical distribution of globular clusters, isolated stars, and dark matter that surrounds a galaxy.

Hawking process The formation of real particles from virtual ones just outside a black hole's event horizon; the means by which black holes evaporate.

head-tail source A radio galaxy whose radio emission is deflected from the galaxy.

heliocentric cosmology A theory of the formation and evolution of the solar system with the Sun at the center.

helioseismology The study of vibrations of the solar surface.

helium flash The explosive ignition of helium fusion in the core of an intermediate-mass, giant star.

helium shell flash The explosive ignition of helium fusion in a thin shell surrounding the core of an intermediate-mass star.

helium shell fusion Helium fusion that occurs in a thin shell surrounding the core of a star.

Herbig-Haro objects Gases heated by outflow from young T Tauri stars.

Hertzsprung-Russell (H-R) diagram A plot of the absolute magnitude or luminosity of stars versus their surface temperatures or spectral classes.

homogeneity The property of the universe being smooth or uniform as measured over suitably large distance intervals.

horizon problem *See* isotropy problem.

horizontal branch stars A group of post-helium-flash stars near the main sequence on the Hertzsprung-Russell diagram.

hot-spot volcanism The creation of volcanoes on a planet's surface caused by a reservoir of hot magma in the planet's mantle under a thin part of the crust.

Hubble classification A system of classifying galaxies according to their appearance into one of four broad categories: spirals, barred spirals, ellipticals, and irregulars.

Hubble constant (H_0) The constant of proportionality in the relation between the recessional velocities of remote galaxies and their distances; the value determines the age of the universe.

Hubble flow The recession of superclusters of galaxies caused by the expansion of the universe.

Hubble law The relationship that states that the redshifts of remote galaxies are directly proportional to their distances from Earth.

hydrocarbon A molecule based on hydrogen and carbon.

hydrogen envelope An extremely large, tenuous sphere of hydrogen gas surrounding the head of a comet.

hydrogen fusion The thermonuclear fusion of hydrogen to produce helium.

hydrogen shell fusion Hydrogen fusion that occurs in a thin shell surrounding the core of a star.

hydrostatic equilibrium A balance between the weight of a layer in a star and the pressure that supports it.

hyperbola An open curve obtained by cutting a cone with a plane.

impact breccia A rock consisting of various fragments cemented together by the impact of a meteoroid.

impact crater A crater on the surface of a planet, moon, or other body produced by the impact of a smaller object.

inferior conjunction The configuration when Mercury or Venus is directly between the Sun and Earth.

inferior planet A planet orbiting closer to the Sun than Earth, namely, Mercury and Venus.

inflation A sudden and extreme expansion of space that occurred once when the universe was very young.

inflationary epoch A brief period shortly after the Big Bang during which the scale of the universe increased much more rapidly than it does today.

infrared radiation Electromagnetic radiation of a wavelength longer than visible light but shorter than radio waves.

instability strip A region on the Hertzsprung-Russell diagram occupied by pulsating stars.

interferometry A method of increasing resolving power by combining electromagnetic radiation obtained by two or more telescopes.

intergalactic gas Gas located between the galaxies within a cluster of galaxies.

intermediate-mass black hole A black hole with a mass between a few hundred and few thousand solar masses.

interstellar extinction The dimming of starlight as it passes through the interstellar medium.

interstellar medium Interstellar gas and dust.

interstellar reddening The reddening of starlight passing through the interstellar medium resulting from the scattering of short-wavelength light more than long-wavelength light.

inverse-square law The gravitational attraction between two objects and the apparent brightness of a light source are both inversely proportional to the square of its distance.

ionosphere (thermosphere) Region of Earth's atmosphere, above the mesosphere, in which sunlight ionizes many atoms.

irregular cluster (of galaxies) An unevenly distributed group of galaxies bound together by their mutual gravitational attraction.

irregular galaxy An asymmetrical galaxy having neither spiral arms nor an elliptical shape.

isotopes Atoms that all have the same number of protons (atomic number) but different numbers of neutrons. Their nuclear properties often differ greatly.

isotropy The fact that the average number of galaxies at different distances from Earth is the same in all directions; also, the fact that the temperature of the cosmic microwave background is essentially the same in all directions.

isotropy problem (horizon problem) The difficulty in explaining why seemingly disconnected regions of the universe have the same temperature.

Jeans instability The condition under which gravitational forces overcome thermal forces to cause part of an interstellar cloud to collapse and form stars and planets.

Jeans unstable Any volume of gas and dust that meets the conditions of a Jeans instability and hence will collapse under the influence of its own gravitational force.

Kepler's laws Three statements, formulated by Johannes Kepler, that describe the motions of the planets.

Kerr black hole Any rotating, uncharged black hole.

kinetic energy The energy an object has as a result of its motion.

Kirchhoff's laws Three statements formulated by Gustav Kirchhoff describing what physical conditions produce each type of spectra.

Kirkwood gaps Gaps in the spacing of asteroid orbits discovered by Daniel Kirkwood and caused by gravitational attractions of planets.

Kuiper belt A doughnut-shaped ring of space around the Sun beyond Pluto that contains many frozen comet bodies, some of which are occasionally deflected toward the inner solar system.

Kuiper belt object (KBO) Space debris orbiting the Sun beyond Neptune in the Kuiper belt. The debris includes small solar system bodies like Pluto, as well as myriad small rocky and icy objects, including many comet nuclei.

law of equal areas (Kepler's second law) The physical law that a line joining a planet and the Sun sweeps out equal areas in equal intervals of time.

law of inertia (Newton's first law of motion) The physical law that an object will stay at rest or move at a constant speed in a fixed direction unless acted upon by a net external force.

law of universal gravitation Newton's law of gravitation, which describes how the gravitational force between two bodies depends on their masses and separation.

lenticular galaxy A disk-shaped galaxy without spiral arms.

light curve A graph that displays variations in the brightness of a star or other astronomical object over time.

light-gathering power A measure of how much light a telescope intercepts and brings to a focus.

light-year (ly) The distance that light travels in a vacuum in 1 year.

lighthouse model The explanation that a pulsar pulses by rotating and funneling energy outward via magnetic fields that are not aligned with the rotation axis.

limb (of the Sun) The apparent edge of the Sun as seen in the sky.

limb darkening The phenomenon whereby the Sun is darker near its limb than near the center of its disk.

line of nodes The line along which the plane of the Moon's orbit intersects the plane of the ecliptic.

liquid metallic hydrogen A metallike form of hydrogen that is produced under extreme pressure.

Local Group The cluster of about 40 galaxies of which our own Galaxy is a member.

long-period comet A comet that takes tens of thousands of years or more to orbit the Sun once.

luminosity The rate at which electromagnetic radiation is emitted from a star or other object.

luminosity class The classification of a star of a given spectral type according to its luminosity and density; the classes are supergiant, bright giant, giant, subgiant, and main sequence.

lunar eclipse An eclipse during which Earth blocks light that would have struck the Moon.

lunar phases The names given to the apparent shapes of the Moon as seen from Earth.

magnetar Especially hot, rapidly rotating neutron stars whose motion helps generate extra-strong magnetic fields.

magnetic dynamo A theory that explains phenomena of the solar cycle as a result of periodic winding and unwinding of the Sun's magnetic field in the solar atmosphere.

magnification The number of times larger in angular diameter an object appears through a telescope than when it is seen by the naked eye.

main sequence A grouping of stars on the Hertzsprung-Russell diagram extending diagonally across the graph from the hottest, brightest stars to the dimmest, coolest stars.

main-sequence star A star, fusing hydrogen to helium in its core, whose surface temperature and luminosity place it on the main sequence on the Hertzsprung-Russell diagram.

mantle That portion of a terrestrial planet located between its crust and core.

mare (*plural maria*) Latin for "sea"; a large, relatively crater-free plain on the Moon.

mare basalt Dark, solidified lava that covers the lunar maria.

mass A measure of the total amount of material in an object.

mascons Regions of relatively high-density matter located not far below the surface of the Moon.

mass–luminosity relationship The direct relationship between the masses and luminosities of main-sequence stars.

matter-dominated universe A universe in which the radiation field that fills all space is unable to prevent the existence of neutral atoms.

mesosphere The layer in Earth's atmosphere above the stratosphere.

metals All elements except hydrogen and helium.

meteor The streak of light seen when any space debris vaporizes in Earth's atmosphere; a "shooting star."

meteor shower Frequent meteors that seem to originate from a common point in the sky.

meteorite A fragment of space debris that has survived passage through Earth's atmosphere.

meteoroid A piece of interplanetary debris smaller than about 10 m (or 10 yd) that is composed primarily of rock and metal.

microlensing The gravitational focusing of light from a distant star or galaxy by a closer object to give a brighter image of the star or galaxy.

Milky Way Galaxy The galaxy in which our solar system resides.

missing mass *See dark matter.*

model A hypothesis that has withstood observational or experimental tests.

molecular clouds Nebulae that are often embedded in much larger bodies of gas and dust.

moment of inertia A measure of how much energy is required to cause an object to rotate.

momentum (more precisely, linear momentum) is a measure of how much energy is available to an object as a result of its motion.

moons (natural satellites) Bodies that orbit larger objects, which in turn orbit stars.

neap tide The least change from high to low tide during a day; it occurs during the first and third quarter phases of the Moon.

nebula (plural nebulae) A cloud of interstellar gas and dust.

neutrino A subatomic particle, with no electric charge and little mass, that is important in many nuclear reactions and in supernovae.

neutron degeneracy pressure A powerful pressure produced by neutrons crowded very close together.

neutron star A very compact, dense stellar remnant composed almost entirely of neutrons.

Newton's laws of motion Newton's equations that describe the motion of matter as a result of forces acting on it.

Newtonian reflector An optical arrangement in a reflecting telescope in which a small, flat mirror reflects converging light rays to a focus on one side of the telescope tube.

Nice model A theory that describes the formation of the planets and other objects orbiting the Sun, with Jupiter and the other giant planets forming first, followed by the inner planets, asteroid belt, Kuiper belt, and Oort comet cloud.

north celestial pole The location on the celestial sphere directly above Earth's northern rotation pole.

northern lights (aurora borealis) Light radiated by atoms and ions in Earth's upper atmosphere due to high-energy particles from the Sun and seen mostly in the northern polar regions.

northern waste (northern devastation) Relatively young, crater-free terrain in the northern hemisphere of Mars.

nova (plural novae) A star in a binary system that experiences a sudden outburst of radiant energy, temporarily increasing its luminosity by a factor of between 10^4 and 10^6 .

nucleosynthesis The formation, by fusion, of higher-mass elements from lower-mass ones.

nucleus (of a comet) A collection of ices and dust that constitutes the solid part of a comet.

OB association An unbound group of very young, massive stars predominantly of spectral types O and B.

OBAFGKM sequence The sequence of stellar spectral classifications from hottest to coolest stars.

objective lens The principal lens of a refracting telescope.

Occam's razor The principle of choosing the simplest scientific theory that correctly explains any phenomenon.

occultation The eclipsing of an astronomical object other than the Moon or Sun by another astronomical body.

Oort cloud A hollow spherical region of the solar system beyond the Kuiper belt where most comets are believed to spend most of their time.

open cluster A loosely bound group of young stars in the disk of a spiral galaxy; a galactic cluster.

open universe A universe with a hyperbolic shape (that is, saddle-shaped); it lacks the mass necessary to someday stop expanding and recollapse. It will expand forever.

opposition The configuration of a planet when it is at an elongation of 180° and thus appears opposite the Sun in the sky.

optical double A pair of stars that appear to be near each other but are unbound and at very different distances from Earth.

orbital inclination The tilt or angle of an object's orbital plane around the Sun compared to the ecliptic.

organic molecule A carbon-based compound.

overcontact binary A close binary system in which the two stars share a common atmosphere.

ozone layer The lower stratosphere, where most of the ozone, O_3 , in the air exists.

pair production The creation of a particle and an antiparticle from energetic photons.

parabola An open curve formed by cutting a circular cone at an angle parallel to the sides of the cone.

parallax The apparent displacement of an object relative to more distant objects caused by viewing it from different locations.

parsec (pc) A unit of distance equal to 3.26 light-years.

partial eclipse A lunar or solar eclipse in which the eclipsed object does not appear completely covered.

Pauli exclusion principle A principle of quantum mechanics that states that two identical particles cannot simultaneously have the same position and momentum.

peculiar galaxy (pec) Any Hubble class of galaxy that appears to be blowing apart.

penumbra The portion of a shadow in which only part of the light source is covered by the shadowmaking body.

penumbral eclipse A lunar eclipse in which the Moon passes only through Earth's penumbra.

perihelion The point in its orbit where a planet or other body in the solar system is nearest the Sun.

period–luminosity relationship A relationship between the period and average luminosity of a pulsating star.

periodic table A listing of the chemical elements according to their properties; created by Dmitri Mendeleev.

photodisintegration The breakup of nuclei in the core of a massive star due to the effects of energetic gamma rays.

photometry The measurement of light intensities.

photon A discrete unit of electromagnetic energy.

photosphere The region in the solar atmosphere from which most of the visible light escapes into space.

pixel A contraction of the term “picture element”; usually refers to one square of a grid into which the light-sensitive component of a charge-coupled device is divided.

plage A bright spot on the Sun believed to be associated with an emerging magnetic field.

Planck era Time from the Big Bang until the Planck time (10^{-43} s).

Planck time The earliest time, about 10^{-43} s after the Big Bang, for which science has equations describing the universe. At that time all four forces in nature today (gravity, electromagnetism, weak, and strong) behaved as one force.

Planck's law A relationship between the energy carried by a photon and its wavelength.

planet An object orbiting a star that is held together by its own gravitational force in a nearly spherical shape, that is able to clear its neighborhood of debris, and is not the moon (or satellite) of a larger orbiting body.

planetary differentiation The process early in the life of each planet whereby denser elements sank inward and lighter ones rose.

planetary nebula A luminous shell of gas ejected from an old, low-mass star.

planetesimal Primordial asteroidlike object from which the planets accreted.

plasma A hot, ionized gas.

plate tectonics The motions of large segments (plates) of Earth's surface caused by convective motions in the underlying mantle.

polymer A long molecule composed of many smaller molecules.

poor cluster (of galaxies) A cluster of galaxies with only a few members.

Population I star A star, such as the Sun, whose spectrum exhibits spectral lines of many elements heavier than helium; a metal-rich star.

Population II star A star whose spectrum exhibits comparatively few spectral lines of elements heavier than helium; a metal-poor star.

positron An electron with a positive rather than negative electric charge; an antielectron.

potential energy The energy stored in an object as a result of its location in space.

powers of ten Representation of numbers using a number between 1 and 9.9999 times 10 to a power.

pre–main-sequence star The stage of star formation just before the main sequence; it involves slow contraction of the young star.

precession (of Earth) A slow, conical motion of Earth's axis of rotation caused by the gravitational pull of the Moon and Sun on Earth's equatorial bulge.

precession of the equinoxes The slow westward motion of the equinoxes along the ecliptic because of Earth's precession.

primary cosmic rays (cosmic rays) High-speed particles traveling through space.

primary mirror The large, concave, light-gathering mirror in a reflecting telescope, analogous to the objective lens on a refracting telescope.

prime focus The point in a reflecting telescope where the primary mirror focuses light.

primordial black hole A relatively low-mass black hole hypothetically formed at the beginning of the universe.

primordial fireball The extremely hot gas that filled the universe immediately following the Big Bang.

primordial nucleosynthesis The transformation by fusion of protons and electrons into hydrogen isotopes, helium, and some lithium in the first few minutes of the existence of the universe.

prograde motion Motion of planet from left to right (eastward) along the celestial sphere against the background stars.

prograde orbit An orbit of a moon or satellite around a planet that is in the same direction as the planet's rotation.

prominence An arch-like concentration of gas seen near the limb of the Sun and extending into the solar corona. The side view of a filament.

proper motion The change in the location of a star on the celestial sphere.

protoplanetary disk (proplyd) A disk of material encircling a protostar or a newborn star.

protostar The earliest stage of a star's life before fusion commences and when gas is rapidly falling onto it.

protosun The Sun prior to the time when hydrogen fusion began in its core.

pulsar A pulsating source generated by a rapidly rotating neutron star with an off-axis magnetic field.

quantum mechanics The branch of physics dealing with the structure and behavior of atoms and their interactions with each other and with light.

quark A particle that is a building block of the heavy nuclear particles such as protons and neutrons.

quasar (quasi-stellar radio source) A starlike object with a very large redshift.

quasi-stellar object (QSO) A quasar.

quintessence One of the explanations of the dark energy causing the universe to accelerate outward.

radial velocity That portion of an object's velocity parallel to the line of sight from Earth.

radial-velocity curve A plot showing the variation of radial velocity with time as seen from Earth for a binary star or variable star.

radiation (photon) pressure The transfer of momentum carried by radiation to an object on which the radiation falls.

radiation-dominated universe The time at the beginning of the universe when the electromagnetic radiation prevented ions and electrons from combining to make neutral atoms.

radiative zone A region inside a star where energy is transported outward by the movement of photons through a gas from a hot location to a cooler one.

radio galaxy A galaxy that emits an unusually large amount of radio waves.

radio lobes Vast regions of radio emission on opposite sides of a radio galaxy.

radio telescope A telescope designed to detect radio waves.

radio wave Long-wavelength electromagnetic radiation.

radioactive Unstable atomic nuclei that naturally decompose by spontaneously emitting particles.

red dwarf A low-mass main-sequence star.

red giant A large, cool star of high luminosity.

redshift The shifting to longer wavelengths of the light from remote galaxies and quasars; the Doppler shift of light from any receding source.

reflecting telescope (reflector) A telescope in which the principal light-gathering component is a concave mirror.

reflection The rebounding of visible light or other electromagnetic radiation off a smooth surface.

reflection nebula A comparatively dense cloud of gas and dust in interstellar space that is illuminated by a star between it and Earth.

refracting telescope (refractor) A telescope in which the principal light-gathering component is a lens.

refraction The bending of light rays when they pass from one transparent medium to another.

refractor See **refracting telescope**.

regolith The powdery, lifeless material on the surface of a moon, asteroid, or planet.

regular cluster (of galaxies) An evenly distributed group of galaxies bound together by mutual gravitational attraction.

resolution See **angular resolution**.

resonance The large response of an object to a small periodic gravitational tug from another object.

retrograde motion The occasional backward (that is, westward) apparent motion of a planet against the background stars as seen from Earth. Retrograde motion is an optical illusion.

retrograde orbit The orbit of a moon or satellite around a planet that is in the direction opposite to the planet's rotation.

retrograde rotation The rotation of a planet opposite to its direction of revolution around the Sun. Only Uranus and Venus have retrograde rotation.

revolution The orbit of one body about another.

rich cluster (of galaxies) A cluster of galaxies with many members.

right ascension (R.A.) The celestial coordinate analogous to longitude on Earth and measured around the celestial equator from the vernal equinox.

rille A winding crack or depression in the lunar surface caused by the collapse of a solidified lava tube.

ringlet Any one of numerous, closely spaced, thin bands of particles in planetary ring systems.

Roche limit The closest distance from a planet or other very massive object that a second, smaller object can remain held together by its own gravitational forces.

Roche lobe The teardrop-shaped regions around each star in a binary star system inside of which gas is gravitationally bound to that star.

rotation The spinning of a body about an axis passing through it.

rotation curve (of a galaxy) A graph showing how the orbital speed of material in a galaxy depends on the distance from the galaxy's center.

RR Lyrae variable A type of pulsating star with a period less than 1 day.

Sagittarius A The strong radio source associated with the nucleus of the Milky Way Galaxy.

scarp A cliff on the Moon or Mercury believed to have formed when those bodies cooled and shrank.

Schmidt corrector plate A specially shaped lens used with spherical mirrors that corrects for spherical aberration and provides an especially wide field of view.

Schwarzschild black hole Any nonrotating, uncharged black hole.

Schwarzschild radius The distance from the center to the event horizon in any black hole.

scientific method The method of doing science based on observation, experimentation, and the formulation of hypotheses (theories) that can be tested.

scientific notation The style of writing large and small numbers using powers of ten.

scientific theory An idea about the natural world that is subject to verification and refinement.

seafloor spreading The process whereby magma upwelling along rifts in the ocean floor causes adjacent segments of Earth's crust to separate.

secondary cosmic rays (cosmic ray shower) Particles from Earth's atmosphere given high speeds Earthward by cosmic rays from space.

secondary mirror A relatively small mirror used in reflecting telescopes to guide the light out the side or bottom of the telescope.

seeing disk The size that a star appears to have on a photographic or charge-coupled-device image as a result of the star's apparent motion smearing the image; it is caused by the changing refraction of the starlight passing through Earth's atmosphere.

seismic waves Vibrations traveling through or along the surface of an astronomical body usually associated with impacts or earthquakelike phenomena.

seismograph A device used to record and measure seismic waves, such as those produced by earthquakes.

semidetached binary A close binary system in which one star fills or is overflowing its Roche lobe.

semimajor axis (of an ellipse) Half of the longest dimension of an ellipse.

SETI The search for extraterrestrial intelligence.

Seyfert galaxy A spiral galaxy with a bright nucleus whose spectrum exhibits emission lines.

shepherd satellite (moon) A small satellite whose gravitational tug is responsible for maintaining a sharply defined ring of matter around a planet such as Saturn or Uranus.

short-period comet A comet that orbits the Sun in the vicinity of the planets, thereby reappearing with tails every 200 years or less.

sidereal month The period of the Moon's revolution about Earth measured with respect to the Moon's location among the stars; 27.3 Earth days.

sidereal period The orbital period of one object about another measured with respect to the stars.

singularity A place of infinite curvature of spacetime; a place (black holes) or time (Big Bang) where our present laws of physics break down.

small solar system bodies (SSSBs) All objects in the solar system that are not planets, dwarf planets, or moons.

snow line The distance from the Sun beyond which ices stayed frozen in the early solar system.

solar corona The Sun's outer atmosphere.

solar cycle A 22-year cycle during which the Sun's magnetic field reverses its polarity twice.

solar day From noontime to the next noontime; for Earth it is 24 hours.

solar eclipse An eclipse during which the Moon blocks the Sun.

solar flare A violent eruption on the Sun's surface.

solar luminosity ($L_{\odot}$) The total energy emitted by the Sun each second.

solar model A set of equations that describe the internal structure and energy generation of the Sun.

solar nebula The cloud of gas and dust from which the Sun and the rest of the solar system formed.

solar system The Sun, planets, their satellites, asteroids, comets, and all other objects that orbit the Sun.

solar wind An outward flow of particles (mostly electrons and protons) from the Sun.

south celestial pole The location on the celestial sphere directly above Earth's south rotation pole.

southern highlands Older, cratered terrain in the Martian southern hemisphere.

southern lights (aurora australis) Light radiated by atoms and ions in Earth's upper atmosphere due to high-energy particles from the Sun; seen mostly in the southern polar regions.

spacetime The concept from special relativity that space and time are both essential in describing the position, motion, and action of any object or event.

spectral analysis Study of the chemistry and motion of objects by examining their spectra.

spectral type A classification of stars according to the appearance of their spectra.

spectrograph A device for photographing a spectrum.

spectroscope A device for directly viewing a spectrum.

spectroscopic binary A double star whose binary nature can be deduced from the periodic Doppler shifting of lines in its spectrum.

spectroscopic parallax A method of determining a star's distance from Earth by measuring its surface temperature, luminosity, and apparent magnitude.

spectroscopy The separation of the wavelengths of light and other electromagnetic radiation emitted by an object.

spectrum (*plural spectra*) The result of electromagnetic radiation passing through a prism or grating so that different wavelengths are separated.

spherical aberration An optical property whereby different portions of a spherical lens or spherical, concave mirror have slightly different focal lengths, thereby producing a fuzzy image.

spicule A narrow jet of rising gas in the solar chromosphere.

spin (of an electron or proton) A small, well-defined amount of angular momentum possessed by electrons, protons, and other particles.

spiral arms Lanes of interstellar gas, dust, and young stars that wind outward in a plane from the central regions of some galaxies.

spiral density wave A spiral-shaped pressure wave that orbits the disk of a spiral galaxy and induces new star formation.

spiral galaxy A flattened, rotating galaxy with pinwheel-like spiral arms winding outward from the galaxy's nuclear bulge.

spoke A moving dark region of Saturn's rings.

spring tide The greatest daily difference between high tide and low tide, occurring when the Moon is new or full.

stable Lagrange points Locations throughout the solar system where the gravitational forces from the Sun and a planet keep space debris trapped.

standard candle An object whose known luminosity can be used to deduce the distance to a galaxy.

starburst galaxy A galaxy where there is an exceptionally high rate of star formation.

Stefan-Boltzmann law The relationship stating that an object emits energy at a rate proportional to the fourth power of its temperature, in kelvins.

stellar evolution The changes in size, luminosity, temperature, and chemical composition that occur as a star ages.

stellar parallax The apparent shift in a nearby star's position on the celestial sphere resulting from Earth's orbit around the Sun.

stellar spectroscopy The study of the properties of stars encoded in their spectra.

stony meteorite A meteorite composed of rock with very little iron; also called a *stone*.

stony-iron meteorite A meteorite composed of roughly equal amounts of rock and iron.

stratosphere The second layer in Earth's atmosphere, directly above the troposphere.

strong nuclear force The force that binds protons and neutrons together in nuclei.

summer solstice The point on the ecliptic where the Sun is farthest north of the celestial equator; the day with the largest number of daylight hours in the northern hemisphere, around June 21.

sunspot A temporary cool region in the solar photosphere created by protruding magnetic fields.

sunspot maximum The time during the solar cycle when the number of sunspots is greatest.

sunspot minimum The time during the solar cycle when the number of sunspots is lowest (often zero).

supercluster (of galaxies) A gravitationally bound collection of many clusters of galaxies.

super-Earth An extrasolar planet with more mass than Earth, but less than 10 Earth masses.

supergiant A star of very high luminosity.

supergranule A large convective cell in the Sun's chromosphere containing many granules.

superior conjunction The configuration when Mercury or Venus is on the opposite side of the Sun from Earth.

superior planet Planet that orbits the Sun beyond the orbit of Earth, namely, Mars, Jupiter, Saturn, Uranus, and Neptune.

supermassive black hole A black hole whose mass exceeds 1000 solar masses.

supernova remnant A nebula left over after a supernova detonates.

superstring theories A set of theories that hopes to describe the nature of spacetime and matter at a more fundamental level than is presently possible.

synchronous rotation The condition when a moon's rotation rate and revolution rate are equal or when a planet's rotation rate equals its moon's revolution rate.

synchrotron radiation The radiation emitted by charged particles moving through a magnetic field; nonthermal radiation.

synodic month (lunar month) The period of revolution of the Moon with respect to the Sun; the length of one cycle of lunar phases; $29\frac{1}{2}$ Earth days.

synodic period The interval between successive occurrences of the same configuration of a planet as seen from Earth.

T Tauri stars Young, variable pre-main-sequence stars associated with interstellar matter that show erratic changes in luminosity.

terminator The line dividing day and night on the surface of any body orbiting the Sun; the line of sunset or sunrise.

terrestrial planet Any of the planets Mercury, Venus, Earth, or Mars; a planet with a composition and density similar to that of Earth.

Theory of Everything Theories under development that comprehensively explain all four fundamental forces in nature.

theory *See* scientific theory.

theory of general relativity A description of spacetime formulated by Einstein explaining how gravity affects the geometry of space and the flow of time.

theory of special relativity A description of mechanics and electromagnetic theory formulated by Einstein according to which measurements of distance, time, and mass are affected by the observer's motion.

thermonuclear fusion A reaction in which the nuclei of atoms are fused together at a high temperature.

3-to-2 spin-orbit coupling The rotation of Mercury, which makes three complete rotations on its axis for every two complete orbits around the Sun.

total eclipse A solar eclipse during which the Sun is completely hidden by the Moon, or a lunar eclipse during which the Moon is completely immersed in Earth's umbra.

trailing-arm spiral galaxy A galaxy with its spiral arms pointing away from the direction of rotation, characteristic of all but one known spiral galaxy.

trans-Neptunian objects (TNOs) Objects orbiting the Sun beyond the orbit of Neptune.

transition (of an electron) The change in energy and orbit of an electron around an atom or molecule.

transition zone Region between the Sun's chromosphere and corona where the temperature skyrockets to about 1 million K.

transverse velocity The portion of an object's velocity perpendicular to our line of sight to it.

Trojan asteroid One of several asteroids at stable Lagrange points that share Jupiter's orbit about the Sun.

troposphere The lowest level of Earth's atmosphere.

Tully-Fisher relation A correlation between the width of the 21-centimeter line of a spiral galaxy and its absolute magnitude.

turnoff point The location of the brightest main-sequence stars on the Hertzsprung-Russell diagram of a globular cluster.

21-cm radiation Radio emission from a hydrogen atom caused by the flip of the electron's spin orientation.

twinkling The apparent change in a star's brightness, position, or color due to the motion of gases in Earth's atmosphere.

Type I Cepheid A Population I Cepheid variable star found in the disks of spiral galaxies.

Type Ia supernova A supernova occurring after a white dwarf accretes enough mass from a companion star to exceed the Chandrasekhar limit.

Type II Cepheid A Population II Cepheid variable star found in elliptical galaxies and in the halos of disk galaxies that is 1.5 magnitudes dimmer than a Type I Cepheid.

Type II supernova A supernova occurring after a massive star's core is converted to iron.

ultraviolet (UV) radiation Electromagnetic radiation of wavelengths shorter than those of visible light but longer than those of X-rays.

umbra The central, completely dark portion of a shadow.

universal constant of gravitation Denoted by the letter G , the constant used in Newton's law of gravitation to determine the acceleration of objects falling toward other objects under the influence of their mutual gravitational attraction.

Van Allen radiation belts Two flattened, doughnut-shaped regions around Earth where many charged particles (mostly protons and electrons) are trapped by Earth's magnetic field.

variable star A star whose luminosity varies.

velocity A quantity that specifies both direction and speed of an object.

vernal equinox The point on the ecliptic where the Sun crosses the celestial equator from south to north; the beginning of spring, around March 21.

very-long-baseline interferometry (VLBI) A method of connecting widely separated radio telescopes to make observations of very high resolution.

virtual particle A particle and its antiparticle, created simultaneously in pairs and which quickly disappear without a trace.

visual binary A double star in which the two components can be resolved through a telescope.

water hole The part of the electromagnetic spectrum at a few thousand megahertz where there is very little background noise from space.

wavelength (λ) The distance between two successive peaks in a wave.

weak nuclear force A nuclear interaction involved in certain kinds of radioactive decay.

weight The force with which a body presses down on the surface of a world such as Earth.

white dwarf A low-mass stellar remnant composed of carbon and oxygen that has exhausted all its thermonuclear fuel and contracted to a size roughly equal to the size of Earth.

Widmanstätten patterns Crystalline structure seen inside iron meteorites.

Wien's law The relationship that the dominant wavelength of radiation emitted by a blackbody varies inversely with its temperature.

winter solstice The point on the ecliptic where the Sun is farthest south of the celestial equator; fewest hours of daylight in the northern hemisphere, around December 22.

Wolf-Rayet stars Rotating stars of at least $20 M_{\odot}$ with strong magnetic fields and stellar winds.

work Change in the kinetic energy (energy of motion) an object experiences when acted on by a force.

wormhole A hypothetical connecting passage between black holes and other places in the universe.

X-rays Electromagnetic radiation whose wavelength is between that of ultraviolet light and gamma rays.

X-ray burster A neutron star in a binary star system that accretes mass, undergoes thermonuclear fusion on its surface, and therefore emits short bursts of X-rays.

Zeeman effect A splitting of spectral lines in the presence of a magnetic field.

zenith The point on the celestial sphere directly overhead.

zero-age main sequence (ZAMS) The positions of stars on the Hertzsprung-Russell diagram that have just begun to fuse hydrogen in their cores.

zodiac A band of 13 constellations around the sky through which the Sun appears to move throughout the year.

zones Light-colored bands in the cloud cover of the four giant planets.

ANSWERS TO COMPUTATIONAL QUESTIONS

CHAPTER 8

28. 5.0×10^{14} tons

APPENDIX A

3.141×10^9 , 3.1831×10^{-9} ; 27,182,820,000;
0.0000000000367879

APPENDIX B

5527°C, 9980°F; -270°C, -454°F

APPENDIX M

580 K, 55 km, temperature decreases with height; +8,
+9, September 1, September 6; 8 $L_{\odot}$, 6000 K

APPENDIX P

Energy increases 9-fold; energy decreases to $\frac{1}{4}$; 2000 J;
0 J; change its moment of inertia, I , and its angular
velocity ω .

APPENDIX Q

$\frac{1}{8}$ kg; ~ 0.35 kg; $\sim 17\frac{1}{2}$

APPENDIX R

Force from Earth in newtons is 4.4 times your weight
in pounds; Force in pounds is your weight in pounds;
same force in both units as your weight; 9×10^{21} N; It
is one-quarter the present force.

INDEX

Page numbers in *italics> indicate figures.*

- A 0620-00 (black hole), 378
- Abell 39 (planetary nebula), 99
- absolute magnitude, 298–299
 - and apparent magnitude, 298
 - and determining distances, 423, 424
 - and H-R diagram, 303–304
- absorption lines, 84, 85, 91
- absorption line spectrum, 87–88
- acceleration, 43
 - of universe, 444, 463–465
- accretion
 - of inner planets, 104
 - and protostars, 323, 324
 - in solar system formation, 103, 104
- accretion disks
 - beta (β) Lyrae system, 344
 - black holes, 377–381, 437
- acetaldehyde, 472
- acetic acid, 472
- achromatic lens, 66
- action and reaction, law of (Newton's third law), 43–44
- active galactic nuclei (AGN), 378, 430
- active galaxies, 430–433, 436–437
- active optics, 70
- active Sun, 273–282
- actuators, 70
- Adams, John Couch, 45, 226
- adaptive optics, 70–71
- Adrastea (moon), 206, 207
- Advanced Composition Explorer*, satellite, 361
- aerogel, 249
- afterglows (of gamma-ray bursts), 381
- albedo, 147
 - of planets, 128
- Aldebaran (red giant), 301, 304
- Aldrin, Buzz, 148
- Algol (star), 343
- ALH84001 (meteorite), 182, 471
- Allende meteorite, 262, 473
- Alnitak, 319, 321
- Alpha Centauri B (star), 118
- alpha particles, 89, 336
- Alpher, Ralph, 446
- Altair (star), 4
- altazimuth coordinate system, 6
- altitude, 6
- Alvan Clark and Sons, 63
- Alvarez, Luis, 263
- Alvarez, Walter, 263
- Amalthea (moon), 206, 207
- amino acids
 - in carbonaceous chondrites, 261–262
 - in comets, 249
- ammonia, on Jupiter, 193, 195
- ammonia ice, 246
- Amor asteroids, 244
- Andromeda Galaxy (M31), 61, 392, 404, 416, 433–434
 - and Milky Way, 412, 414
 - supernova in, 359
- angle of incidence, 58
- angle of reflection, 58
- angstrom ($\text{\AA}$), 52
- angular diameter, 7
- angular momentum, 43–44, 45
 - conservation of, 102
 - of neutron stars, 364
 - see also* spin
- angular resolution, 60–62
- angular size, 7
- annihilation
 - of most of matter and antimatter in early universe, 452–453
 - particles and antiparticles, 453
- annular eclipse, 21
 - in 2001–2020, 22
- anorthosite, 146–147
- Antares (star), 99, 304
- anthropomorphism, 318
- antimatter, in early universe, 452–453
- antimuons, 288
- antineutrinos, 452
- antiparticles, 451, 452, 453
 - and annihilation of particles, 453
- antiquarks, 452
- antitau particles, 288
- Antlia (dwarf galaxy), 415
- Apennine Mountains, 263
- aphelion, of planet orbit, 38, 39
- Aphrodite Terra (Venus), 166, 167
- Apollo (asteroid), 239
- Apollo 11*, 145, 146, 148, 154
- Apollo 12*, 145
- Apollo 13*, 145
- Apollo 14*, 145, 154
- Apollo 15*, 144, 147, 154
- Apollo 16*, 147, 148
- Apollo 17*, 145, 146
- Apollo asteroids, 244
- Apollo* missions, 77, 145–146, 147, 148, 154
- apparent binaries, *see* optical doubles
- apparent magnitude, 297–298
 - and absolute magnitude, 298
 - determining distances with, 423, 424
- Aquarius (the Water Bearer), 16, 351, 448
- Aquila (the Eagle), 4
- Arabia Terrae (Mars), 168, 171
- arc angle, 7
- arcminutes (arcmin), 7
- arcseconds (arcsec), 7
- Arcturus (star), 4, 304
- Arecibo Observatory, 477, 479
- Arecibo Observatory radio telescope, 73
- Ares Vallis (Mars), 176, 181–182
- argon, 130, 175, 288
- Argyre Planitia (Mars), 177
- Aries (the Ram), 16
- A ring, 210, 215
- Aristarchus of Samos, 31, 32
- Aristillus crater (Moon), 142
- Armstrong, Neil, 145, 148
- asterisms, 2–3
- asteroid 2004 FU162, 244
- asteroid 87 Sylvia, 242
- asteroid belt, 103, 107, 114, 239, 240
 - asteroids orbiting outside, 244–246
 - Jupiter's gravity and gaps in, 241–244
- asteroid collisions, 241, 242–243
- asteroid impacts
 - on Earth, 263–264
 - on Jupiter, 199
 - on Mercury, 157
- asteroid P/2013 P5, 243
- asteroid P/2013 R3, 242
- asteroids, 24, 107, 108, 234, 238–246, 259
 - outside asteroid belt, 244–246
 - Earth-crossing, 244
 - families of, 243
 - iron meteorites, 260
 - outside solar system, 114
 - and stony meteorites, 261
 - Tunguska event, 257
 - see also* dwarf planets; meteoroids
- asteroid-satellite systems, 242
- astrobiology, 469–480
- astrometric method for detecting
 - exoplanets, 116
- astronomical distances, 23, 294–296
 - units of, 40
- astronomical measurement, 128
- astronomical unit (AU), 39, 40
- astronomy, nonoptical, 71–79
- astrophotography, 239
- asymptotic giant branch (AGB) stars, 349
- atomic number, 89
- atoms, 88–90
 - neutral, 90, 455
- A type stars, 303
- Auriga (the Charioteer), 4
- auroras, 140
 - from coronal mass ejection, 282
 - on Ganymede, 205
 - on Jupiter, 197, 198, 199
 - on Saturn, 207, 208
- Australia, asteroid impact in, 264
- autumnal equinox, 6, 11, 12
- average density
 - of matter in early universe, 456
 - of planets, 127, 128
- azimuth, 6
- Baade, Walter, 362, 426
- Babcock, Horace, 278
- Balmer, Johann, 91
- Balmer absorption lines, 91, 302–303

- Balmer series, 91
- Baptistinas (asteroids), 263
- Barnard 86 (nebula), 320
- Barnard's star, 92
 - proper motion of, 93
- barred spiral galaxies, 394–395, 408–409, 411, 412
- Barringer Crater, *see* Meteor Crater
- barycenter, of Earth–Moon, 152
- basalt, 146, 147, 165
- beliefs, theories and, 30
- Bell, Jocelyn, 362
- belt asteroids, 239
- belts and zones
 - on Jupiter, 193, 196
 - on Neptune, 224, 225
 - on outer planets, 228, 229
 - on Saturn, 207–208, 209
 - on Uranus, 220
- beryllium, 336
- Bessel, Friedrich Wilhelm, 294, 352
- beta (β) Lyrae system, 343
- beta (β) Persei, *see* Algol (star)
- Beta Pictoris (star), 114, 115
- Betelgeuse (supergiant), 301, 304, 355
- Biermann, Ludwig, 250, 251
- Big Bang, 98, 444, 446–449, 455, 458
- Big Dipper, 2–3, 7
 - as pointer star, 3
- binary head-tail source, 432
- binary neutron stars, 367–369
- binary pulsars, 367, 374
- binary stars
 - detached, 342, 343
 - eclipsing, 309–310, 343
 - orbital motion of, 311–312
 - orbits of, 373
 - overcontact, 342, 343, 344
 - and stellar masses, 307–310
- binary star systems, 100, 308, 353
 - and black holes, 377–378, 380–381
 - center of mass of, 309
 - close, 310
 - mass transfer in, 340–344
 - spectral-line motion in, 312
 - stellar remnant black holes in, 377–378
 - see also* close binary systems; semidetached binary systems; spectroscopic binaries; visual binaries
- bipolar planetary nebula, 351, 352
- birth line, 325
- blackbodies, 80–81
- blackbody curves, 80, 81, 82
 - and stars, 300
- blackbody radiation, 79–83
- blackbody spectrum, 85, 87
- black hole collisions, 382
- black holes, 24, 369, 373–381, 437
 - characteristics of, 374
 - effect of on time, 377
 - evaporation of, 383
 - evidence for, 377–381
 - gases near, 435
 - and gas jets, 380–381
 - gravitational tidal force of, 383
 - infalling matter, 376
 - matter in, 373
 - tidal force of, 376
 - types of, 375
 - see also* stellar black holes; supermassive black holes
- blazars, 432
- BL Lacertae (BL Lac) objects, 432, 433, 437
 - central engine, 437
- blueshift, 54
- Bohr, Niels, 302
- Bok, Bart, 322
- Bok globules, 322
 - protostar in, 323
- bolide, 256–257
- Boltzmann, Ludwig, 82, 83
 - see also* Stefan-Boltzmann law
- Bondi, Herman, 447
- BOOMERANG, 446, 447
- Boötes (the Herdsman), 4, 382, 422
- boron, and existence of life, 471
- Brahe, Tycho, 36, 37, 294, 347, 359
- Braille (asteroid), 256
- Braille, Louis, 256
- B ring, 210–211
- Brooks, Grace, 302
- brown dwarfs, 325
- brown ovals, 193
- B stars, 303, 327, 328, 398, 406, 459
 - see also* OB associations
- buckyballs, 472–473
- Bunsen burner, 84
- Bunsen, Robert, 84–85
- Burns Cliff (Mars), 179
- Burst and Transient Source Experiment (BATSE), 381
- butterfly diagram, of sunspots, 276
- Caesar, Julius, 13–14
- caldera, 174
- Callisto (moon), 181, 205, 206, 474
- calorimeters, 77
- Caloris Basin (Mercury), 155–157, 158
- Camelopardus (the Giraffe), 92
- Canada-France-Hawaii Telescope (CFHT), 67
- Canes Venatici constellation, 411
- Canis Major (the Large Dog), 4
- Canis Major Dwarf Galaxy, 400, 415
- Canis Minor (the Smaller Dog), 4
- Cannon, Annie Jump, 302
- carbon
 - creation from beryllium, 336
 - and existence of life, 471, 472–473
 - helium fusion and, 98
- carbonaceous chondrites, 472, 473
- carbon dioxide
 - on Earth, 130, 131, 132–133
 - on Mars, 175
 - on Venus, 163, 164
- carbon dioxide ice (dry ice), 181, 246
- carbon fusion, 354, 355, 356
- carbon monoxide
 - frozen, 237
 - interstellar gas, 321
 - on Pluto, 237
- Carina constellation, 334
- Carpenter, Alta, 302
- Cartwheel Galaxy, 416
- Cassegrain, Laurent, 58
- Cassegrain focus, 53, 58, 59, 86
- Cassegrain mirrors, 66
- Cassini, Giovanni, 194, 210
- Cassini division, 210–212
- Cassini spacecraft, 194, 196, 198, 200
 - Enceladus, 219
 - Rhea, 218
 - Saturn, 208, 209
 - Saturn's rings, 211, 212, 213, 214
 - Titan, 217, 218
- Cassiopeia A (Cas A), 357, 358, 359, 366, 426
- Castalia (asteroid), 241
- Castelli, Benedetto, 274
- cause-and-effect link, sunspots and
 - temperatures, 279
- celestial equator, 5, 6, 10, 11, 16
- celestial sphere, 5–6
- centaur objects, 246
- Centaurus constellation, 294, 395, 431
- Centaurus X–3, 368
- center of mass, 308–309
- central bulge
 - of barred spirals, 408–409
 - in Milky Way, 396
 - in spiral galaxies, 403–405
- central peaks (in craters), 142
- Cepheid variable stars, 391–394, 424
- Cerenkov, Pavel A., 289
- Cerenkov flashes, 360
- Cerenkov radiation, 288–289
- Ceres (asteroid), 107
- Ceres (dwarf planet), 108, 234, 237, 239, 240
- CERN particle accelerator, 449
- cesium, 85
- Challenger* space shuttle, 27
- Chandrasekhar, Subrahmanyan, 77, 353
- Chandrasekhar limit, 353–354
- Chandra X-ray Observatory, 77, 79
 - black holes, 380
 - Coma cluster, 416
 - galactic nucleus, 398
 - galaxy 3C 294, 458
 - intergalactic gas, 416, 420
 - NGC 6240, 419
 - novae, 353
 - supermassive black holes, 436
 - X-ray emission, 431
- Chandrayaan–1*, 148
- Chaney, Dave, 123
- charge-coupled devices (CCDs), 67–68, 75, 86, 300
 - infrared-detecting, 75
- Charon (moon), 235–237
- Chelyabinsk, Russia, bolide over, 256, 257
- Cherenkov radiation, 78
- Chicxulub Crater, Mexico, 263, 264
- Chihuahua, Mexico, Allende meteorite, 262
- chlorofluorocarbons (CFCs), 57, 132
- chocolate chip cake analogy for expanding
 - universe, 425
- chondrites, 261, 472
- chondrules, 261, 262
- Christy, James, 236
- chromatic aberration, 64, 65
- chromosphere, 269–270, 279–280
- circumpolar stars, 8

- circumstellar shell, 355
 classical Kuiper belt objects, 246, 247
Clementine spacecraft, 148, 149
 close binary systems, 310
 white dwarfs in, 353–354
 closed universe, 462
 clusters (of galaxies), 412–421
 and dark matter, 460
 distribution of, 414
 motion of, 422
 redshifts and blueshifts of, 464
 cobalt
 creation of, 357
 nickel decay to, 354
 collimators, 78
 collision-ejection theory of Moon formation, 150, 151
 colors
 of light, 50, 51
 and temperatures, 79–83
 wavelengths of, 51, 52, 53
 coma (of a comet), 249, 251–252
 Coma Berenices constellation, 391, 415
 Coma clusters (of galaxies), 415, 420
 Comet 2006 SQ372, 254
 Comet 67P/Churyumov-Gerasimenko, 252, 253
 Comet Brooks, 252
 comet dust, 249, 254
 Comet Hale-Bopp, 233, 247, 250, 254–255
 Comet Halley, 45, 248, 254, 472
 Comet Holmes, 249
 Comet Hyakutake, 247
 Comet Ikeya-Seki, 252
 comet impacts, 252–253
 on Jupiter, 198–199
 Comet LINEAR (Lincoln Near Earth Asteroid program), 246, 254
 Comet Mrkos B, 251
 comet nuclei, 108, 115, 247–248
 and comet tails, 250
 formation of solar system, 108
 comets, 24, 105, 108, 234, 239, 246–256
 breakup of, 254
 formation of, 246
 and formation of life, 472
 Kuiper belt, 105, 106
 mass loss, 254–255
 orbits and tails of, 250
 outside solar system, 114
 structure of, 251
 sungrazing, 255, 256
 Comet Schwassmann-Wachmann 3, 254, 255
 Comet Shoemaker-Levy 9, 198, 199, 254
 comet tails, 249–252
 Comet Tempel 1, 252, 253, 254
 Comet Temple-Tuttle, 258
 Comet Toit-Neujmin-Delporte, 254
 Comet West, 24, 250
 Comet Wild 2, 248–249, 472
 compact disks (CDs), 86
 comparative planetology, 126–129
 of the inner planets, 184–185
 comparison spectrum, 86
 complex molecules, and existence of life requirements, 471–472
 compression waves, 407
 Compton, Arthur Holly, 77
 Compton Gamma-Ray Observatory, 77, 381
 computer-aided analysis, 100, 102
 concave lenses, 63
 conceptual framework, 32
 Cone Nebula, 99, 100, 329
 configuration, 33–35
 confinement period, 452
 conic sections, 45
 conjunction of planets, 35
 conservation of angular momentum, 43–44, 102
 constellations, 3–5
 circumpolar, 8
 cycle of, 9–10
 constructive interference, 52
 contact binaries, 342, 343
 continental drift, 134
 continuous line spectrum, 87, 88
 continuum, 428
 convection
 on Earth, 137, 138
 solar granulation and, 269, 270
 on Sun, 283, 286
 convection current, 137, 138
 convective zone (Sun), 283, 286
 convergent boundaries, 134–135
 convex lenses, 63
 Copernicus, Nicolaus, 32–36
 Cordelia (moon), 222
 core bounce, 356
 core collapse, 356
 core helium fusion, 333, 335
 globular cluster stars, 339
 intermediate-mass stars, 349
 core hydrogen fusion, 333–334, 348
 Corona Borealis, 422
 coronagraph, 272
 coronal holes, 280
 coronal loops, 278
 coronal mass ejection, 140, 281–282
 coronal spectrum, 271
Cosmic Background Explorer (COBE), 74, 446, 447
 Milky Way, 397
 temperature differences, 456
 cosmic censorship, 377
 cosmic light horizon, 452
 cosmic microwave background, 446–448, 456, 462
 and curvature of space, 462
 motion through, 448
 sound waves in, 456
 temperature differences in, 456, 457
 cosmic microwave background photons, 463
 cosmic microwave background radiation, 448, 455
 cosmic neutrino background, 447
 cosmic neutrinos, 447
 cosmic rays, 273, 361
 cosmic ray shower, 361
 cosmological constant, 444, 464–465
 cosmological redshift, 445, 446
 cosmology, 30, 443–468
 defined, 444
 coudé focus, 59, 86
 Crab Nebula, 99, 364, 365
 Crab pulsar, 364, 365
 craters, 107–108
 on Earth, 143, 257
 on Mars, 171, 173, 177
 on Mercury, 155–158, 159
 on Moon, 141–143, 144, 145
 C ring, 211
 cryovolcanoes, 228
 Curie, Marie, 88
Curiosity rover, 169, 177–179, 182, 183
 Curtis, Heber D., 391, 434
 Cushman, Florence, 302
 CW Leonis (star), 255
 Cygnus (the Swan), 4, 5, 323
 Cygnus (planetary nebula), 351
 Cygnus A (Cyg A, 3C 405), 426, 427, 432
 Cygnus Loop, 322, 358
 Cygnus X-1 system, 378

 Dactyl (asteroid), 107, 242
 Daphnis (moon), 212
 dark ages, 456, 458
 dark energy, 461, 463–465
 dark matter, 402–403, 419–421, 456, 461
 early universe, 463
 formation of, 452
 galaxy structures and, 412
 Milky Way, 396
 dark matter halos, 419, 420
 dark nebulae, 319, 321, 323
 Barnard, 320
 dark side of the Moon, 16–17
 Darwin, Sir George, 154
 Davis Raymond, 288
 day, on Mercury, 160, 161
 day-night cycle, 8–10
 debris, in solar system, 105–108
 declination (Dec.), 5, 6
 decoupling, 454–455
Deep Impact spacecraft, 252, 253
 deferent, 32, 33
 degenerate electrons, 335
 degenerate matter, 353
 Deimos (moon), 183–184
 Dells (Mars), 179
 Delphinus (the Dolphin), 4
 Delta Aquarids meteor shower, 258
 delta (δ) Cephei stars, *see* Type I Cepheid variables
 Deneb (star), 4
 dense cores, 100
 Bok globules, 322
 density, 99
 and evolution of universe, 454
 of planets, 127, 128
De revolutionibus orbium coelestium (*On the Revolutions of the Celestial Spheres*), 35–36
De stella nova (*On the New Star*), 36
 destructive interference, 52
 detached binaries, 342, 343
 deuterium, 77
 and neutrino detectors, 288
 Dicke, Robert, 446
 differential rotation
 Jupiter, 196
 Jupiter's clouds, 194

- differential rotation (*cont.*)
 - Milky Way, 401, 406
 - Saturn, 207–208
 - Sun, 275, 277, 278
- diffraction gratings, 85–87
- diffraction gratings, 86, 87
- dinosaur extinction, 263–264
- Dione (moon), 216
- Dirac, Paul A.M., 456
- direct imaging method of detecting
 - exoplanets, 116–117
- direct motion, of planets, 30
- distance–magnitude relationship, 299, 338, 423, 424
- distance-measuring techniques, 423, 424
- distance-velocity graph, 422, 423
- diurnal motion, 8
- divergent boundaries, 134–135
- Doppler, Christian, 54
- Doppler effect, 54
 - and clusters of galaxies, 445
 - cosmic microwave background, 448
- Doppler shift, 54
 - and discovery of exoplanets, 116
 - interstellar reddening and, 320
 - radial velocity and, 92
 - solar granules and, 269
 - special relativity, 370
 - in spectral lines, 311
 - speeds of stars and, 93
- Dorado (the Goldfish), 430
- double-line spectroscopic binary, 311, 312
- double pulsar, 367
- double-radio sources, 431, 432, 436, 437
- double star, 307, 308
- Draco (the Dragon), 14
- Drake, Frank, 123, 471, 477
- Drake equation, 477–478
- dry ice (carbon dioxide ice), 246
 - on Mars, 181
- dust devils, on Mars, 176–177
- dust tail (of a comet), 251–252
- DVDs, 86
- dwarf elliptical galaxies, 410, 414
- dwarf planets, 108, 234–238
 - in outer solar system, 246
 - see also* Ceres (dwarf planet); Eris (dwarf planet); Haumea (dwarf planet); Make-make (dwarf planet); Pluto
- dynamo theory (of Earth's magnetic field), 138
- Dysnomia (moon), 237, 238

- E0 to E7 galaxies, 410
- Eagle Nebula, 327
- early type galaxies, 409
- Earth, 125, 129–141
 - albedo of, 128
 - asteroid impacts, 263–264
 - atmosphere, 68–71, 130–133, 185, 473
 - atmospheric temperature, 131
 - average density of, 128
 - axis of rotation, 10, 11, 14–16, 150, 152
 - core of, 135–137
 - craters on, 143, 257
 - crust, 133–135, 137
 - ecliptic, 6–7, 9, 10, 11
 - and electromagnetic radiation, 56, 57
 - formation of, 104–105
 - greenhouse effect, 130
 - infalling space debris, 259, 260
 - infrared radiation, 132
 - interior of, 135–137, 159, 185
 - iridium in, 260
 - lunar eclipses, 18–20
 - magnetic field of, 137–141, 185, 223
 - magnetosphere, 139
 - mantle of, 135, 136–137
 - mass of, 127, 185
 - oceans, 130
 - orbital data, 126
 - pieces of the Moon and Mars on, 259
 - planetesimal impact on, 105
 - plant life on, 130
 - precession and axis of rotation, 14–16
 - revolution of, 9–10, 13
 - rotation of, 8, 13–14, 185
 - seasons on, 10–16
 - shadow of, 18, 19
 - size of, 126, 127, 185
 - surface of, 185
 - synodic and sidereal periods, 35
 - temperatures, 136, 137
 - vital statistics, 129
 - water on, 287
 - weather, 131, 132–133
 - see also* Moon
- Earth-centered cosmology, *see* geocentric cosmology
- Earth-crossing asteroids, 244
- earthly cycles, 8–10
- Earth–Moon distance, 21
- Earth–Moon system, 152–153
- earthquakes, 134, 137
- Earth–Sun distance, 11, 23, 126
- eclipse, 18–22
- eclipsing binaries, 311, 343
 - light curves of, 309–310
- ecliptic, 6–7, 9, 10
 - line of nodes and, 18–19
- Eddington, Arthur, 283
- Einstein, Albert, 282
 - cosmological constant of, 444, 464
 - on expanding universe, 461
 - general relativity theory, 30, 117, 367, 371–374
 - on light, 54–55, 83
 - relativity theories, 369–374
 - special relativity theory, 369–371
- Einstein cross, 437, 438
- Einstein Observatory
 - NGC 5128, 431
- Einstein ring, 437
- ejecta blanket, 142
- electric charge, of electrons and protons, 89
- electromagnetic force, 89
- electromagnetic radiation, 50, 52, 53
 - intensity of, 55
 - properties of, 83
 - types of, 55–57
 - as waves, 52
 - see also* X rays; gamma rays; infrared radiation; radio waves; ultraviolet (UV) radiation; visible light
- electromagnetic spectrum, 55, 56, 78, 80
- electromagnetism, 88
 - unification of with other forces, 449, 450
- electron degeneracy pressure, 335
 - in high-mass stars, 356
 - in intermediate-mass stars, 351
- electron neutrinos, 288
- electrons, 88–92
 - degenerate, 335
 - early universe, 452, 453
 - excited state of, 90–91
 - formation of, 451
 - orbits of, 89, 90
- electron spin, 394
- electroweak force, 449, 451
- elementary particles, in early universe, 452–453
- elements
 - defined, 84
 - identifying with spectral analysis, 84–88
 - periodic table of, 85, 86
 - stars' transformation of matter from lighter to heavier, 98
- ellipse, 37–38
- elliptical galaxies, 409–411, 412
 - early universe, 459–460
 - spectra of, 422
- elongation (of a planet), 33–35, 38
- $E = mc^2$, 282, 370, 383, 453
- emission lines, 84, 85
- emission line spectrum, 87
- emission nebulae, 319, 321, 327, 395
- Enceladus (moon), 219–220
 - water in, 110
- Encke, Johann Franz, 212
- Encke division, 212
- energy, creation of matter from, 453
- epicycle, 32–33
- Epimetheus (moon), 213
- EPOXI spacecraft, 248
- Epsilon (ε) Eridani, 114, 471
- equations of nuclear fusion, 332
- equations of stellar structure, 325
- equatorial bulge, 14, 15
- equinoxes
 - precession of, 16
 - and seasons, 10–11, 12
- era of recombination, 455
- ergoregions, 375, 376
- E ring, 220
- Eris (dwarf planet), 108, 234, 237, 238
- Eros (asteroid), 24, 245
- Eta Aquarids meteor shower, 258
- ethane, on Titan, 218
- ethyl alcohol, 472
- Eugenia, 242
- Eurasian plate, 134
- Europa (moon), 181, 202–204, 474
 - water in, 110
- Europa torus, 197, 207
- European Space Agency, 77, 252, 296
- evening star, 33
 - Venus as, 162
- event horizon, 374, 376–377, 383
- evolutionary tracks (of stars), 325
- excited state, 90–91
- exoplanets, 113–124, 471, 474, 478
 - classifications, 119
 - methods of detecting, 114–118
 - number of, 122
 - orbits, 120
 - search for life on, 123

- separations between their stars and, 121
- sizes, masses, and compositions, 118
- experiments, 29
- Explorer 42* satellite, *see* *Uburu* satellite
- extrasolar planets, *see* exoplanets
- extraterrestrial biology, 472
- extraterrestrial-generated radio signals,
 - search for, 123
- Extreme Ultraviolet Explorer (EUVE), 77
- eyepiece lenses, 58, 62, 63, 64
- Eyjafjallajökull volcano (Iceland), lightning
 - generated by, 164
- faint young sun paradox, 287
- fall (autumn), 11
- falls (meteorite impacts), 262
- false-color techniques, 74
- false vacuum, 451
- far side of the Moon, 16–17, 145, 149, 151–153, 156
- Far Ultraviolet Spectroscopic Explorer (FUSE), 77
- Fermi, Enrico, 288
- Fermi Large Area Telescope, 77
- filaments
 - Milky Way, 398
 - of solar corona, 279, 280, 281
- finds (meteorites), 262
- fireballs, 256–257
- first quarter Moon, 16, 17
- Fisher, Richard, 424
- Fleming, Williamina, 302
- flocculent spiral galaxies, 404, 405, 406–407
- focal length, 58, 59, 62, 63, 64
- focal plane, 58, 64
- focal point, 58, 59
- focus (*plural* foci), of ellipse, 37
- Fomalhaut (star), 115, 117
- Fomalhaut b, 116–117, 120
- force, 43
- force law (Newton's second law), 43
- formaldehyde, 472
- forsterite, 249
- 4U 1543-47 (black hole), 378
- frame of reference, 36
- Fraunhofer, Joseph von, 84, 85
- free-floating planets, 120
- frequency, 54–55
 - of photon, 54
- Friedmann, Alexandre, 461
- F ring, 212, 213, 214
- From the Earth to the Moon* (Verne), 141
- frost line, *see* snow line
- F type stars, 303
- fullerenes, 263
- full Moon, 16, 17, 18
 - apparent magnitude, 297
- fusion, 98–99
 - see also* core helium fusion; core hydrogen fusion; helium fusion; helium shell fusion; hydrogen fusion; hydrogen shell fusion; neon fusion; nuclear fusion; oxygen fusion; silicon fusion; thermonuclear fusion
- fusion crust, of meteorites, 260
- galactic cannibalism, 400, 418, 419
- galactic cosmic rays, 273
- galactic disk, 402
 - of Milky Way, 393–397, 399–400
- Galactic Evolution Explorer (GALEX), 77
- galactic habitable zone, 475
- galactic mergers, 418
- galactic nucleus (Milky Way), 392, 393, 396, 397–399, 426
- galaxies, 24, 391
 - active, 430–433, 436–437
 - clusters of, 412–421
 - colliding, 418
 - distances to and redshifts, 422–425
 - E0 to E7, 410
 - early type, 409
 - early universe, 459–460
 - formation of, 455–459, 464
 - Hubble classification of, 403
 - Hubble tuning fork diagram of, 411–412
 - interacting, 418
 - merging, 418, 419
 - properties of, 412
 - with rings, 416
 - superclusters of, 412–413
 - types of, 403–412
 - see also* elliptical galaxies; giant elliptical galaxies; grand-design spiral galaxies; irregular galaxies; lenticular galaxies; peculiar galaxies; spiral galaxies
- galaxy collisions, 412, 416–419, 460
- Galilean moons, 41, 42, 199–205
 - vital statistics, 200
- Galilei, Galileo, 41–42, 199, 274, 275
- Galileo Regio (Ganymede), 204
- Galileo spacecraft
 - and Callisto, 205
 - destruction of, 204
 - and Europa, 202, 203–204
 - and Io, 202
 - and Jupiter, 195, 197, 198, 200, 201
- Galle, Johann, 45, 226
- Galle crater (Mars), 173
- gamma-ray bursts, 381–384
- gamma-ray photons, and
 - photodisintegration, 356
- gamma rays, 55, 56, 77, 78, 80
 - early universe, 446, 453
 - and giant stars, 336
 - positron-electron collisions, 399
 - and Sun, 110, 282, 283, 284, 285, 287
- gamma-ray telescopes, 77–78
- Gamow, George, 446
- Ganymede (moon), 181, 204–205, 474
 - size of, 126
 - water in, 110
- gas giants, 104, 126, 127, 128, 195
- Gaspra (asteroid), 107
- gas tail (of a comet), 251–252
- Gemini (the Twins), 4
- Geminids meteor shower, 258
- general relativity, equations of, 375–376
- general relativity, theory of, 367, 371–374, 420, 454
 - black holes and, 374, 376
 - expanding universe and, 444, 461
 - gravitational lensing and, 437
 - gravitational redshift and, 376
- Genesis* spacecraft, 250, 273
- geocentric cosmology, 30–31, 32–33
- geodesics, 372
- giant branch, 349
- giant elliptical galaxies, 409, 410, 412, 418
- giant molecular clouds, 319, 321–323, 395
 - H II regions, 328
- giant phase (of stars), 333–334, 339
- giant planets
 - formation of, 103–104
 - see also* outer planets
- giant stars, 304, 305
 - evolution of, 335–336
 - and helium flash, 335
 - mass loss of, 334
 - as variables, 337
- Gill, Edith, 302
- Gill, Mabel, 302
- Giotto* spacecraft, 248, 472
- Gliese 229A, 325
- Gliese 229B, 325
- Gliese 581, 120
- Gliese 581 b, c, and e, 120
- Gliese 581d, 471
- Gliese 876, 119
- Gliese 876d, 119
- glitch (of pulsars), 366–367
- Global Positioning System (GPS), 30
- global warming, 11, 132–133
- globular clusters, 118, 338–340
 - black holes in, 379
 - formation of, 416
 - Milky Way, 393, 399, 400, 401
- glucose, 471
- glycine, 472
- Gold, Thomas, 447
- goldilocks zone, 122
- Gondwana, 135
- gossamer rings, Jupiter's moons, 207
- graben, 148, 149
- grand-design spiral galaxies, 404, 405, 406–407, 408
- Grand Unified Theories (GUTs), 449, 450
- Gran Telescopio Canarias, 71
- granulation, 269, 270
- granules, in photosphere, 269, 270
- grating spectrograph, 86
- gravitation, 14, 15, 27, 89, 371
 - at all scales, 47
 - inverse square law for, 44
 - and other physical forces, 449, 450
 - solar system formation and, 99–100
- gravitation, Newton's law of universal, 42, 44–46, 372
- gravitational attraction, and general relativity, 371
- gravitational lensing, 420, 421
 - of quasars, 437–438
- gravitational microlensing, 117, 118, 119
- gravitational radiation, 374
- gravitational redshift, 372–373, 376
- gravitational wave detectors, 374
- gravitational waves, 374
- gravity, *see* gravitation
- gravity, repulsive, 464
- Great Attractor, 448
- Great Dark Spot (Neptune), 226
- Great Dying, 263, 264
- greatest eastern and greatest western elongations, 33

- Great Nebula of Orion, *see* Orion Nebula
 Great Observatories, NASA, 70, 75, 76, 77
 see also Chandra X-ray Observatory; Compton Gamma-Ray Observatory; Hubble Space Telescope; Spitzer Space Telescope
 Great Red Spot (Jupiter), 193–194, 195
 Greek astronomers, ancient, 30, 32
 star classification, 296
 greenhouse effect, 287
 on Earth, 130, 132–133
 on Venus, 164–165
 Gregorian calendar, 14
 Gregory XIII, Pope, 14
 G ring, 214
 GRO J042+32, 378
 GRO J1655–40, 378
 ground state, of electrons, 90
 GS 1124–T68 (black hole), 378
 GS 2000+25 (black hole), 378
 G type stars, 303
 Gulf of Suez, 134
 gullies, Martian, 180
 Gum Nebula, 358, 365
- H 1705–250 (black hole), 378
 habitable zone, 122, 474, 475
 Hadley's rille (Moon), 144
 Hale, George Ellery, 276, 278
 Hall, Asaph, 184
 Halley, Edmond, 45
 halo field stars, 399
 halo system, Milky Way, 396, 399–400
 Harvard Observatory, 302
 Harwood, Margaret, 302
 Haumea (dwarf planet), 108, 234, 237
 Hawking, Stephen, 379–380, 383
 Hawking process, 383
 Hawking radiation, 383
Hayabusa space probe, 245
 H $_{\alpha}$ (H alpha) Balmer line, 91
 HD 148937, 355
 HD 149026, 119
 HD 171978 (binary), 312
 HD 209458, 119
 HD 209458b, 119
 HD 82943, 120
 HD85512b (planet), 471
 HDE 26868 (star), 377
 head-tail sources, 432
 Heat Shield Rock, 179
 heliocentric cosmology, 31–36, 41–42
 helioseismology, 277–278, 282
 heliosphere, 272–273
 helium, 85
 Cepheids, 337
 in early universe, 454
 on Earth, 130
 hydrogen fusion in Sun creating, 283, 285, 287
 interstellar medium, 319
 on Jupiter, 194, 195, 196
 in red dwarfs, 332–333
 on Saturn, 210
 in solar system formation, 98–99, 100, 103
 helium flash, 335, 348, 349
 helium fusion, 98, 334–335, 348, 349, 356
 giant stars, 335
 neutron stars, 369
 see also core helium fusion
 helium nuclei, 336
 helium rain, 210
 helium shell flash, 350
 helium shell fusion, 348, 349
 Helix Nebula (NGC 7293), 351
 Hellas Planitia (Mars), 171
 Helmholtz, Hermann von, 198
 hematite blueberries, 177, 179
Henry Draper Catalogue, 119, 312
 Herbig-Haro objects, 325, 326
 Hercules cluster of galaxies, 412, 413, 416
 Hercules X–1, 368
 Herschel, William, 45, 55, 175, 226, 237, 307, 350, 392, 393
 hertz, 54–55
 Hertz, Heinrich, 54, 55
 Hertzsprung, Ejnar, 303–304
 Hertzsprung-Russell (H-R) diagrams, 303–305
 absolute magnitude, 303–304
 giant branch, 348
 globular cluster, 339
 and luminosity classes, 306
 Pleiades, 320
 plotting star clusters on, 328–330
 pre-main-sequence stars, 325
 structure of, 340–341
 white dwarfs, 352
 zero-age main sequence (ZAMS), 330
 H'iaka (moon), 237
 Higgs, Peter, 29
 Higgs boson, 29
High-Energy Transient Explorer, 382
 H II regions, 326–328
 barred spiral galaxies, 395
 Himalayas, 134, 136
Hinode satellite, 270, 281
 HIP 77900 b (exoplanet), 120
 Hipparchus, 16, 296
Hipparcos satellite (High Precision Parallax Collecting Satellite), 296, 339
 Hirayama, Kiyotsugu, 243
 Homestake neutrino detector, 288
 homogeneity, 449, 451–452
 Hooke, Robert, 28–29, 194
 horizon, 2
 horizon problem, *see* isotropy problem
 horizontal branch stars, 339, 349
 Horologium constellation, 417
 Horsehead Nebula, 321
 hot spots
 cosmic microwave background, 462
 on Jupiter, 195
 X-ray sources, 368
 hot-spot volcanism, 174
 Hourglass Nebula, 351, 352
 Hoyle, Sir Fred, 447
 H $_{\alpha}$ (H alpha) photons, 91
 Huang, Su-Shu, 344
 Hubble, Edwin, 391–392, 409, 410
 on expanding universe, 445
 on galaxy motion, 422
 Hubble classification of galaxies, 403
 Hubble constant, 423, 424, 428, 445–446
 Hubble Deep Field images
 elliptical galaxies, 460
 Hubble flow, 422, 424, 444, 452
 Hubble law, 423, 425, 428
 Hubble Space Telescope, 69, 70, 71, 116, 119, 297
 black holes, 378
 colors of stars, 300
 dark matter distribution, 460
 discovering asteroids, 240
 distant galaxies, 425–426
 galaxy interactions, 412
 isolated stars, 418
 M32 and M87, 434
 Near Infrared Camera and Multi-Object Spectrometer, 75–76
 Neptune, 226
 NGC 4261, 435
 novae, 353
 Pluto and its moons, 236
 protogalaxies, 458
 radio lobes, 427
 Saturn, 207
 Saturn's rings, 213
 SN 1997ff, 463
 Uranus, 220, 222
 Hubble tuning fork diagram, 411–412
 Hulse, Russell, 374
 Humason, Milton, 422
 Huygens, Christiaan, 50, 169, 171, 215
Huygens probe, 217, 218
 hybrid gamma-ray burst, 381
 Hydra (moon), 236
 Hydra constellation, 117, 422
 hydrocarbons, on Titan, 216, 218
 hydrogen
 atomic structure of, 90
 dark ages of universe, 456
 early universe, 454
 on Earth, 130
 emission spectrum of, 87
 energy level diagram of, 90
 interstellar medium, 319
 ionized, 90
 on Jupiter, 194, 195, 196
 in red dwarfs, 332
 on Saturn, 207
 in solar system formation, 98–99, 100, 103
 spectrum of, 87
 in star atmospheres, 301
 hydrogen, interstellar, 394
 hydrogen, neutral, 90
 hydrogen atom, spin flipping, 394
 hydrogen burning, *see* hydrogen fusion
 hydrogen cyanide, on Titan, 216
 hydrogen envelope, 249, 250
 hydrogen fusion, 98, 283, 284–285, 287, 356
 neutron stars, 369
 pre-main-sequence star, 324–325
 in Sun's development, 110
 see also core hydrogen fusion
 hydrogen isotopes, 90
 hydrogen molecules, in interstellar medium, 321
 hydrogen peroxide, on Mars, 183
 hydrogen shell fusion, 333, 348, 349
 globular cluster stars, 338, 339
 hydrogen sulfide, Venus, 163

- hydrostatic equilibrium, 283
- hydroxyl radical, 148
- hyperbolas, 45
- Hyperion (moon), 192, 215, 216
- hypernova, 382
- hyperthermophiles, 476
- hypervelocity stars, 401
- hypotheses, 29

- Iapetus (moon), 213, 216
- IC 2163, 418
- IC 2220, *see* Toby Jug (star)
- Icarus (asteroid), 239
- ice, 204
 - on Ceres, 237
 - of comets, 246, 248, 249, 254
 - on Enceladus, 219, 220
 - on Europa, 202, 203
 - on Ganymede, 204
 - on Haumea and Namaka, 237
 - on Jupiter, 198
 - on Mars, 171, 175, 181
 - on Mercury, 160, 161
 - on Miranda, 224, 225
 - on Neptune, 224, 225, 227
 - on outer planets, 192
 - on Pluto and Charon, 236, 237
 - on Rhea, 218
 - in Saturn's rings, 212, 213
 - see also* ammonia ice; carbon dioxide ice (dry ice); methane ice
- ice giants, 104, 126, 127, 128, 220
- Iceland, 134
- Ida (asteroid), 107, 242
- ideal blackbody curve, 83
- impact breccias (Moon), 147
- impact craters
 - on Callisto, 205
 - from comets, 252
 - on Earth, 257
 - on Mercury, 158
 - from meteors, 257
- impact rate, early solar system, 107, 108
- impacts, of meteorites, 257
- Index Catalogs (IC), 321
- Indian-Australian plate, 134
- inertia, law of, 42–43
- inferior conjunction, 33
- inferior planets, 33
- inflation (universe expansion rate), 451–452, 462
- inflationary epoch (of universe), 450, 451–452, 455–456
- Infrared Astronomical Satellite (IRAS), 73, 75, 398
- infrared-detecting CCDs, 75
- infrared radiation, 55, 56, 57, 75–77, 80
 - on Earth, 132
- Infrared Space Observatory (ISO), 75
- Infrared Telescope Facility (IRTF), 75
- infrared telescopes, 75–77
- inner planets
 - comparative planetology of, 184–185
 - comparison of, 185
 - formation of, 104–105
 - see also* Earth; Mars; Mercury; Venus
- instability strip, 336–337
- intensity, and temperature, 82
- intensity-versus-wavelength curve, 87
- interference patterns, 52
- interferometry, 71
 - and radio images, 73–74
- intergalactic clouds of gas, 457–458
- intergalactic gas, 24, 416, 420
- intermediate-mass black holes, 378, 379, 399
- International Astronomical Union, 108
- International Space Station, 43, 45
- International Ultraviolet Explorer (IUE), 77
- interplanetary debris, 109
- Interstellar Boundary Explorer (IBEX), 273
- interstellar clouds
 - collisions, 322–324
 - galactic center, 399
 - Milky Way, 397–398
- interstellar extinction, 320
- interstellar gas and dust, 24, 293, 318–322
 - black holes, 379
 - emission spectra from, 92
 - and galactic structure, 407
 - galaxy collisions, 416
 - and mapping Milky Way, 394
 - Milky Way, 358, 397
 - molecular clouds, 321–322
 - solar nebula, 100–102
 - and spiral density waves, 406, 407
 - and star formation, 99–100, 407
 - supernovae, 356
- interstellar medium, 318–322
 - planetary nebula, 351
 - and supernovae, 358
- interstellar reddening, 320, 321
- intracluster gas, 436
- inverse-square law, 44, 298, 299
- Io (moon), 201–202
- ionic bonding, 471
- ionization, 90
 - and photons, 454
- ionosphere, 132
- ion tail (of a comet), 251–252
- Io torus, 197, 202, 207
- iridium, 260, 263
- iron
 - on Earth, 136–137
 - ionized, 90
 - in Mercury's core, 158–160
 - spectral lines of, 85
- iron core, of supergiants, 355
- iron meteorites, 260–261
- iron oxides
 - on Earth, 131
 - on Mars, 175
- irons, 260
- irregular clusters of galaxies, 415
- irregular galaxies, 411, 412
- Irvine-Michigan-Brookhaven (IMB) detector, 360
- Ishtar Terra (Venus), 166
- island universe theory, 391
- isotopes, 90
- isotropy, 448–449, 451–452
- isotropy problem, 451
- Itokawa (asteroid), 241, 245

- James Webb Space Telescope, 76, 123
- Jansky, Karl, 72, 74, 426
- JAXA, 245
- Jeans, Sir James, 100
- Jeans instability, 100, 321, 322
- John Paul II, Pope, 42
- joule, 54
- joule-seconds, 54
- Jovian planets, 128
 - see also* Jupiter; Neptune; Saturn; Uranus
- Julian calendar, 13–14
- Juno (asteroid), 239
- Juno spacecraft, 198
- Jupiter, 24, 192–207
 - albedo of, 128
 - and asteroids, 239
 - atmosphere of, 193, 195–196, 229
 - aurorae, 197, 198, 199
 - average density of, 128, 194
 - belts and zones on, 193, 196
 - chemical composition of, 193, 194–196
 - clouds of, 193, 194, 195–196
 - comet and asteroid impacts, 198–199
 - core of, 198
 - differential rotation, 196
 - formation of, 103
 - gravity of, and asteroids, 241–244
 - ice on, 198
 - interior of, 196–198, 228, 229
 - magnetic field of, 223, 228, 229
 - magnetosphere, 197, 198
 - mantle, 196
 - mass of, 127, 193, 194, 195, 228
 - moons of, 42, 129
 - orbital data, 126
 - rings of, 206–207
 - rotation, 228
 - and Saturn, 105
 - size of, 126, 127, 193, 228
 - synodic and sidereal periods, 35
 - temperatures on, 195–196, 198
 - vital statistics, 192
 - winds on, 196
 - see also* Galilean moons
- Jupiter–Sun distance, 35, 40, 126

- Kamiokande II neutrino detector, 288
 - and SN 1987A, 360
- Kant, Immanuel, 391
- kappa (κ) Arietis, 311, 312
- Karin (asteroid), 243
- Karl G. Jansky Very Large Array, 74
- KBO 1993 SC, 247
- KBO 1998 WW31, 247
- Keck I and Keck II telescopes, 62, 70, 71, 72, 297
 - distant galaxies, 425
 - protogalaxies, 458
 - Titan, 218
- Keeler, James, 212
- Keeler gap, 212
- Keenan, P.C., 305
- Kelvin, Lord (William Thomson), 198
- Kelvin-Helmholtz mechanism, 198
- kelvins, 81
- Kennedy, John F., 145
- Kepler, Johannes, 28, 29, 36, 359
- Kepler 9, 120
- Kepler 37, 119, 120
- Kepler 37b, 119
- Kepler 70b, 120

- Kepler 186f, 122
- Kepler's laws, 28, 29, 37–42
- Kepler's second law (law of equal areas), 39
- Kepler's third law, 39–41, 399, 402, 435
 - Newton's derivation of, 40–41, 42, 307–308
 - and orbital speed, 420, 435
 - and stellar masses, 307, 308, 309
- Kerberos (moon), 236
- Kerr, Roy, 375
- Kerr black holes, 375, 376, 377, 435
- kilolight years, 40
- Kin Peak National Observatory, 78
- Kirchhoff, Gustav, 84–85, 87
- Kirchhoff's laws, 87–88
- Kirkwood, Daniel, 241
- Kirkwood gaps, 241
- Kitt Peak National Observatory, 8
- Kormendy, John, 434
- Kreutz, Heinrich, 255
- Kreutz comets, 255
- K type stars, 303
- Kuiper, Gerard, 105
- Kuiper belt, 105, 106, 234, 237
 - comets, 246–248
- Kuiper belt objects (KBOs), 105, 108, 247

- L1014 dark nebula, 323
- Lacerta (the Lizard), 432
- Lagrange, Joseph, 244
- lakes, frozen, on Triton, 227–228
- Lakshmi Planum (Venus), 166
- λ_{max} , 80
- Land of the Midnight Sun, 12, 13
- Large Binocular Telescope II, 67
- large impact theory, *see* collision-ejection theory of Moon formation
- Large Magellanic Cloud, 411
 - supernova 1987A, 359
- Laser Interferometer Gravitational Wave Observatory (LIGO), 374
- Late Heavy Bombardment, 107–108, 143, 145, 148, 150
 - Mercury, 158
- late type galaxies, 409
- Laurasia, 135
- lava
 - on Io, 202
 - on Mercury, 155, 156, 157, 158
 - on Moon, 145, 147, 150–152
 - on Venus, 165, 166
- lava tubes, on Moon, 144, 145
- law of action and reaction (Newton's third law), 43–44
- law of equal areas (Kepler's second law), 39
- law of inertia (Newton's first law), 42–43
- leading-arm spiral galaxies, 409
- leap years, 13–14
- Leavitt, Henrietta, 391
- Leland, Evelyn, 302
- LeMaitre, Georges, 446, 461
- length contraction, 370
- lenses, 63, 66
 - see also* eyepiece lenses; refracting telescopes
- lenticulae, on Europa, 203–204
- lenticular galaxies, 403, 409, 411
- Leo (the Lion), 4
 - cosmic microwave background, 448
- Leo I galaxy, 410
- Leonid meteor shower, 258, 259
- Lepus (the Hare), 244, 325
- Leverrier, Urbain-Jean-Joseph, 45, 226
- Levy, David, 198, 246
- LGM1 (Little Green Men 1), 362
- librations, lunar, 152
- Lick Observatory, 63
- life
 - existence requirements, 471–474
 - search for on Mars, 182–183
 - on Titan, 218
- light
 - colors of, 50, 51
 - and gravitational curving of space, 372
 - nature of, 50–57
 - particle properties of, 50, 54
 - and spacetime, 372–373
 - wave nature of, 50–52, 54
 - from white dwarfs, 373
- light, speed of, 53–54, 371
 - and special relativity, 369–370
- light curves
 - of binaries, 309–310
 - of novae, 353
 - of supernova, 354
- light-gathering power (telescopes), 60
- lightning
 - generated by Eyjafjallajökull volcano in Iceland, 164
 - lightning storms in Jupiter's magnetosphere, 197
 - on Saturn, 208
- light pollution, 68–70
- light rays, reflecting telescopes, 57–60
- light-year (ly), 40
- limb (of solar disk), 269
- limb darkening, 269
- Lin, C.C., 407
- Lindblad, Bertil, 401, 406
- line of nodes, 18–19
- liquid metallic hydrogen
 - on Jupiter, 198
 - on Saturn, 210, 215
- liquid water, and search for extraterrestrial life, 470, 471, 474
- lithium, 120
 - early universe, 454
 - in solar system formation, 98
- lithosphere, 137
- Little Ice Age, 279
- LMC X-3 (black hole), 377
- Local Arm, 396
- Local Group, 414, 415
- Local Supercluster, 413
- long-period comets, 254
- loop quantum gravity, 449
- Lowell, Percival, 169
- luminosity, 298–299
 - and H-R diagram, 303–304
 - quasars, 429–430
 - and stellar spectra, 305–306
 - see also* mass–luminosity relation; period–luminosity relation
- luminosity classes, on H-R diagrams, 306
- Lunar Crater Observing and Sensing Satellite (LCROSS), 148

- lunar eclipses, 18–20
 - in 2015–2018, 20
- lunar highlands, 145, 146, 147, 150
- lunar month, *see* synodic month
- lunar phases, 16–18
- Lunar Prospector*, 147, 148, 149
- lunar ranging, 154
- Lunar Reconnaissance Orbiter*, 144, 148, 149
- Lyman-alpha forest, 429
- Lyman series, 91
- Lynden-Bell, Donald, 433
- Lyra (the Lyre), 4, 14, 114
- Lyrids meteor shower, 258

- M10 (globular cluster), 338
- M13 (globular cluster), 478, 479
- M15 (globular cluster), 351
- M16 (nebula), *see* Eagle Nebula
- M32 (elliptical galaxy), 404, 433–434
- M33 (galaxy), 405
- M43, 113
- M55 (cluster), 339
- M58 (galaxy), 400
- M74 (galaxy), 405
- M81 (galaxy), 403, 417
- M82 (galaxy), 379, 417
- M83 (galaxy), 395
- M84 (galaxy), 409
- M86 (galaxy), 409
- M87 (giant elliptical galaxy), 432, 434
 - black hole in, 378, 381
- M100 (galaxy), 391, 403
- M103 (star cluster), 334
- M104 (galaxy), 404
- M105 (galaxy), 410
- Maat Mons (Venus), 166
- Mab (moon), 224
- McCandless, Bruce, 27
- MACHOs (massive compact halo objects), 402–403
- Mackie, Johanna, 302
- McMath-Pierce Solar Telescope, 78
- Magellan* spacecraft, 165–166, 167
- MAGIC array, 78
- magma, 137
- magnesium, 136
- magnetar, 365–366
- magnetic dynamo model, 278–279
- magnetic fields, 137–141
 - black hole accretion disk, 436
 - flip of, 278, 279
 - of Mars, 171, 173, 174
 - of Mercury, 159
 - of neutron stars, 364–366
 - north and south poles of, 278
 - plages, filaments, and prominences, 279–280, 281
 - strength of, 364
 - of Sun, 272–273, 276–282
 - sunspots, 276–279
 - Venus's apparent lack of, 163, 167–168
- magnetosphere, 138, 139
- magnetotail, Jupiter, 197, 198
- magnification, with telescopes, 62
- magnitude scales, 296–299
- main sequence, 304
 - lifetimes of stars, 332
 - stellar mass, 326

- main-sequence stars, 304, 305–306, 329, 330–332
 early and middle evolution of, 333–336
 and mass–luminosity relation, 310–311
 red dwarfs, 332–333
- Makemake (dwarf planet), 108, 234, 237
- Marconi, Guglielmo, 478
- Mare Australe, 145
- mare basalt, 146, 147
- Mare Imbrium (Sea of Showers), 143, 145, 158, 259
- Mare Ingenii, 145
- Mare Moscoviense, 145
- Mare Nectaris (Sea of Nectar), 143
- Mare Nubium (Sea of Clouds), 143
- Mare Orientale, 145, 158
- Mare Serenitatis (Sea of Serenity), 143, 146
- Mare Tranquillitatis (Sea of Tranquility), 143, 144, 145, 148
- maria (*sing.*, mare), 143–144, 150
- Mariner 2*, 250
- Mariner 9*, 169
- Mariner 10*, 155, 157, 158, 159
- Mars, 168–184
 albedo of, 128
 and asteroids, 241, 244
 atmosphere of, 175–177, 185
 avalanches on, 181
 “canals” on, 169
 chemical composition of, 183
 core of, 174–175
 craters on, 169–172, 173, 177
 crust of, 174
 dust on, 175–177
 formation of, 104–105
 gravitational force of, 175
 hematite on, 179
 ice on, 171, 175, 181
 interior of, 172–175, 185
 life on, 474
 magnetic field of, 185
 magnetic fields of, 171, 173, 174
 mantle of, 174
 mass of, 127, 185
 meteorites from, 181, 182, 183, 259
 methane on, 183
 moons of, 129, 183–184
 northern waste or devastation and southern highlands, 169
 orbital data, 126
 regolith, 182–183
 retrograde motion of, 31–32, 34
 rivers and lakes on, 177, 178
 rotation of, 185
 search for microscopic life on, 182–183
 seasons on, 175, 176
 size of, 127, 185
 surface of, 169–172, 173, 185
 synodic and sidereal periods, 35
 tectonic plate motion on, 174
 temperatures on, 176
 topography of, 170
 vital statistics, 168
 volcanoes, 169, 171, 172, 173–174
 water on, 105, 169, 171, 172, 174, 175, 177–182
 winds and storms on, 176–177
- Mars Global Surveyor*, 170, 172, 173, 178, 179, 180
- Mars meteorite ALH84001, 471
- Mars Pathfinder*, 170, 176
- Marsquakes, 181
- Mars Reconnaissance Orbiter*, 174, 175, 181, 182
- Mars–Sun distance, 35, 126
- mascons (mass concentrations), 148
- mass, 40–41
 center of, 308–309
 and force, 43
 and pre-main-sequence star evolution, 324–326
- mass extinctions, 263–264
- mass loss
 of supergiants, 355
 from supermassive star, 326
- mass–luminosity relation, 310–311
- mass transfer, in binary systems, 340–344
- mathematical models, 304
- Mather, John C., 447
- Mathilde (asteroid), 245, 246, 261
- matter
 in black hole, 374–376
 creation of, from energy, 453
 in early universe, 452–453
- matter-dominated universe, 454
- Mauna Kea, Hawaii, 67, 68, 71, 72
- Maunder, E. Walter, 279
- Maunder minimum, 279
- Maxwell, James Clerk, 51–52, 53, 55
- Maxwell Montes (Venus), 166
- mean solar day, 13
- megahertz, 55
- megalight years, 40
- Menzel 3 (Mz 3), 351
- Mercury, 154–162
 albedo of, 128, 155
 atmosphere of, 161–162, 185
 craters on, 155–158, 159
 elongation of, 33–35
 formation of, 104–105
 interior of, 158, 159, 185
 iron core of, 158–160
 Late Heavy Bombardment, 108
 magnetic field of, 159, 185
 magnetosphere of, 159
 mantle of, 159, 160
 mass of, 127, 185
 maximum elongation of, 38
 and Moon, 156
 orbital data, 126, 160
 orbit of, and general relativity, 373
 outgassing, 161
 rotation and revolution of, 185
 size of, 126, 127, 156, 185
 solar day, 160
 surface of, 154–158, 159, 160, 185
 synodic and sidereal periods, 35
 temperatures on, 156, 161–162
 terminator, 157
 vital statistics, 155
 volcanoes on, 156, 157, 158
- Mercury–Sun distance, 35, 126
- mesosphere, 132
- Messenger* (Mercury Surface, Space, Environment, Geochemistry, and Ranging), 155–159, 160
- Messier Catalogue, 327
- metal-poor stars, *see* Population II stars
- metal-rich stars, *see* Population I stars
- metals, defined, 99
- Meteor Crater, 257, 264
- meteorites, 234, 256, 257, 258–262
 ancient, 472
 water in, 261
- meteoroids, 24, 108, 234, 256–257
 Tunguska event, 257
- meteors, 234, 256, 257–258
- meteor showers, 257–258
- meteorwrongs, 261
- methane
 frozen, 237
 on Jupiter, 195
 on Mars, 183
 on Titan, 216, 218
 on Uranus, 220
- methane ice, 246
 Neptune, 224, 225, 227
- Metis (moon), 206, 207
- Mexico, Chicxulub Crater, 263, 264
- MGC 2237, *see* Rosette Nebula (NGC 2237)
- microcraters, 143
- microlensing, 117, 118, 119, 402–403, 420
- micrometers (mm), 55
- microns (mm), 55
- microwaves, 55, 56
- microwave sky, 448
- microwave telescopes, 74–75
- Mid-Atlantic Ridge, 134
- Milky Way Galaxy, 1, 390–403
 black holes in, 377–378
 central regions, 76
 disk of, 358, 390, 394–397, 399–400, 401, 402
 Drake equation and extraterrestrial civilizations in, 477–478
 galactic center, 393, 397
 gamma-ray burst in, 381
 globular clusters in, 339
 halo of, 390, 402, 403
 Local Group, 414
 luminosity of, 432
 mass of, 402
 microwave sky map, 448
 motion of, 448
 novae, 353
 open clusters in, 340
 planetary nebulae, 351
 rotation of, 400–402
 spiral arms and arm segments, 390, 395, 396, 400
 supernovae in, 358
 supernova remnants in, 358
- Miller, Stanley, 473
- Miller-Urey experiment, 473
- millisecond pulsars, 367
- Mimas (moon), 211, 216
- Minkowski, Rudolph, 426
- Minor Planet Center, 239–240
- minor planets, *see* planetesimals

- Miranda (moon), 224, 225
 mirrors
 focal length of, 58
 in telescopes, 57–60
 see also reflecting telescopes
 missing mass, in Milky Way, 402
 Mizar A and Mizar B (binary system), 308
 models, 29, 31, 161
 Moe (Martian rock), 176
 molecular clouds, 319
 molecules, 90
 momentum, *see* angular momentum
 Monoceros (the Unicorn), 67, 323, 330
 star formation in, 100
 Monoceros giant molecular cloud, 323
 see also Cone Nebula
 month, 17
 see also sidereal month; synodic month
 Moon, 125, 141–154
 albedo of, 147
 atmosphere, 150
 core of, 149
 craters on, 107, 141–143, 144, 145, 150, 152
 creation of, 105
 crust of, 145, 147, 149
 during day, 17
 and Earth's seasons, 14
 far side and dark side of, 16–17, 145, 149, 151–153, 156
 formation of, 150–152
 highlands, 145, 146, 150
 interior of, 147–148, 150
 Late Heavy Bombardment, 107–108
 lava on, 145, 147, 150–152
 librations, 152
 magnetic field of, 150
 mantle of, 149
 maria, 150
 meteorite from, 259
 meteorite impacts on, 258
 movement away from Earth, 154
 phases of, 16–18
 recent impacts on, 259
 regolith, 146, 147
 sidereal month, 17–18
 surface of, 141–145
 synchronous rotation of, 152–153, 154
 synodic month, 17–18
 terminator, 17
 and tides on Earth, 152, 153–154
 umbra and penumbra of, 20, 21
 vital statistics, 141
 see also entries under lunar
 Moon–Earth system, 152–153
 moonlets, in Saturn's rings, 213
 moonquakes, 147
 Moon rocks, 90, 146–147, 150
 dating of, 108
 moons (natural satellites), 108
 of outer planets, 192, 228–229
 of planets, 128–129
 potential for developing and supporting life, 474
 see also shepherd satellites
 Morgan, W.W., 305
 morning star, 33
 Venus as, 162
 motion, *see* direct motion; diurnal motion; planetary motion; retrograde motion, of planets
 Mount Everest, 136
 MS 1054-03 (galaxy cluster), 426
 M stars, 303
 M-theory, 449
 muon neutrinos, 288
 muons, 288

 Namaka (moon), 237
 nanometer (nm), 52
 Nasmyth, James, 59
 Nasmyth focus, 59, 86
 National Aeronautics and Space Agency (NASA)
 astrobiology, 476
 cataloging dangerous asteroids, 264
 Explorer 42, 367
 Great Observatories, 70, 75, 76, 77
 High-Energy Transient Explorer, 382
 IBEX, 273
 and search for extraterrestrial intelligence, 471
 SOHO, 282
 National Radio Astronomy Observatory, 73, 123
 natural satellites, *see* moons (natural satellites)
 neap tides, 154
Near Earth Asteroid Rendezvous (NEAR) Shoemaker spacecraft, 245, 246
 Near Infrared Camera and Multi-Object Spectrometer (NICMOS), 75–76
 nebulae, 318, 391
 see also dark nebulae; emission nebulae; planetary nebulae; reflection nebula; spiral nebulae
 neon, in the Sun, 273
 neon fusion, 355, 356
 Neptune, 224–228
 albedo of, 128
 and asteroids, 244
 atmosphere of, 229
 average density of, 128
 belts and zones on, 224, 225
 clouds, 224
 discovery of, 45, 226
 formation of, 104
 gravity and comets, 246
 Great Dark Spot, 226
 ice on, 224, 225, 227
 images of, 71
 interior of, 222, 226, 228, 229
 magnetic field of, 223, 226, 228, 229
 mass of, 127, 228
 moons of, 129, 227–228
 orbital data, 126
 and Pluto, 105, 235, 247
 rings of, 226–227
 rotation of, 224–226, 228
 and Saturn, 105
 size of, 126, 127, 228
 storms on, 226
 synodic and sidereal periods, 35
 tides on, 227, 228
 trans-Neptunian objects, 106
 vital statistics, 221
 Neptune–Sun distance, 35, 126
 neutral atoms, 90, 455
 neutral hydrogen (H I), 90, 327, 394–395
 neutrino detectors, 288–289
 and Supernova 1987A, 360
 neutrinos, 285, 287–289, 356
 cosmic, 447
 and dark matter, 402
 early universe, 452, 453
 formation of, 451
 neutron degeneracy pressure, 362, 369, 373
 neutrons, 89–90
 early universe, 452, 453
 transformation into protons, 449
 neutron stars, 118, 362–369
 atmosphere of, 366
 axis of rotation of, 364–365
 collision of, 367
 and gas jets, 380–381
 internal structure of, 366–367
 magnetic fields of, 364–366
 mass of, 369
 pulsars, 364–365
 remnants, 373
 rotation of, and magnetic fields, 364–366
 and X-ray bursters, 369
New Astronomy (Kepler), 37, 38
New General Catalog (NGC), 321
New Horizons spacecraft, 207, 237
 new Moon, 16, 18
 New Technology Telescope, 70
 Newton, Isaac, 28–29, 40–41, 43
 on continental drift, 134
 creating parabolic surface, 66
 and Kepler's third law, 40–41, 42, 307–308
 on light, 50
 and static universe, 444
 telescopes, 57–58
 Newtonian reflector, 58, 59
 Newton's first law (law of inertia), 42–43
 Newton's law of universal gravitation, 29, 30, 44–46, 372
 Newton's laws of motion, 42–44, 372
 Newton's second law (force law), 43
 Newton's third law (law of action and reaction), 43–44
 NGC 604, 92
 NGC 664, 424
 NGC 891, 404
 NGC 1265, 432
 NGC 1313, 379
 NGC 1357, 403
 NGC 1512, 417
 NGC 1566, 430
 NGC 2023 (reflection nebula), 321
 NGC 2024 (emission nebula), 321
 NGC 2207, 418
 NGC 2237, *see* Rosette Nebula (NGC 2237)
 NGC 2264, 329–330
 NGC 2363, 92
 NGC 3077, 417
 NGC 3115, 435
 NGC 3377, 410
 NGC 3393, 418
 NGC 3576, 321
 NGC 3603, 321
 NGC 4151, 430

- NGC 4261, 435
 NGC 4406, 410
 NGC 4485, 411
 NGC 4490, 411
 NGC 4676, 418
 NGC 5128 (Centaurus A), 431
 NGC 5907, 404
 NGC 6240, 418, 419
 NGC 6520 (star cluster), 320
 NGC 6826, 351
 NGC 7052, 378, 379
 Nice model, 100, 105, 109, 118–119, 120, 150, 198, 213, 240
 nickel, radioactive decay to cobalt, 354
 night sky, 1, 2–7
 nitrogen
 on Earth, 130, 131
 and existence of life, 471, 472
 frozen, 237
 on Mars, 175
 on Pluto, 237
 on Titan, 216
 Nix (moon), 236
 NN Serpens (binary star), 309
 nonoptical astronomy, 71–79
 Norma (the Carpenter's Square), 351, 355
 Norma Arm, 396
 North American plate, 134, 136
 north and south magnetic poles, 138, 139
 north celestial pole, 5, 6, 14
 precession and path of, 15
 northern lights (aurora borealis), 140
 North Star, *see* Polaris (star)
 nova, 353
 light curve, 353
 Nova Cygni 1975, 353
 Nova Herculis 1934, 353
 nuclear bulge, *see* central bulge
 nuclear density, 356
 nuclear fission, 135
 early universe, 453
 nuclear fusion, 135, 332
 nuclear physics, 88
 nucleosynthesis, 354, 367
 nucleus (of an atom), 88–90

 OBAFGKM sequence, 303
 OB associations, 328, 329, 411
 objective lenses, 63, 64, 65
 observations, 29, 30
 confirming, 240
 Occam's razor, 29, 174, 251, 437
 occultation, 222
 Olbers, Heinrich, 238
 Oliver, Bernard, 474–476
 olivine, 180, 261
 Olympus Mons (Mars), 171, 172, 174
 1RXS 1609 (star), 117
 Oort, Jan, 247
 Oort cloud, 105, 106, 246, 247–248
 open clusters (of stars), 99, 322–323, 328, 329–330, 340
 open universe, 462
 Ophelia (moon), 222
 Ophiuchus (the Serpent Holder), 13
 globular cluster in, 338
 merging galaxies in, 419
Opportunity rover, 170, 176, 177, 179, 180, 181

 opposition (of planets), 35
 optical doubles, 307
 orbital eccentricity, 37–38, 126
 orbital inclination, 109, 126
 orbital period, 126
 orbits
 of exoplanets, 120
 of planets, 108, 109, 126
 shapes of, 45
 O'Reilly, Mollie, 302
 organic molecules, extraterrestrial, 472–474
 Orion (the Hunter), 4
 and asterism, 2, 3
 dark and emission nebulae, 321
 light images, 73
 stars in, 297
 see also Betelgeuse (supergiant)
 Orionids meteor shower, 258
 Orion Molecular Cloud, 99, 113, 114
 Orion Nebula, 100, 102, 113, 114, 321, 328
 Orion's belt, 4
 Orion Spur, 396
 O stars, 303, 327, 328, 397, 406, 459
 Outer Arm, 396
 outer planets, 191–232
 comparison of planetary features, 228–229
 see also Jupiter; Neptune; Saturn; Uranus
 outgassing, 161
 overcontact binaries, 342, 343, 344
 oxygen
 atomic structure of, 89
 carbon and creation of, 336
 on Earth, 130–131
 on Ganymede, 205
 on Mars, 175
 photon emission, 92
 on Venus, 165
 oxygen fusion, 354, 355, 356
 oxygen isotopes, 89
 ozone, 57, 131–132
 ozone holes, 132
 ozone layer, 57, 131–132

 Pacific plate, 134, 136, 174
 pair production, 453
 Pallas (asteroid), 107, 238–239, 240
 pallasite, 261
 Pallene (moon), 213
 Palomar Observatory, 68, 69, 426
 Pan (moon), 212
 Pandora (moon), 212, 213
 Pangaea, 134, 135
 parabolas, orbits, 45
 parallax, 36, 37, 294–296
 to determine distance, 295, 296
 see also spectroscopic parallax; stellar parallax
 parallax angle, 294–296, 299
 Paranal Observatory, 49, 71
 parsec, 40, 296
 Parsons, William, Lord Rosse, 391
 partial eclipse, 19
 partial solar eclipse, 21
 particles, in early universe, 452–453
 Paschen series, 91
Pathfinder, 181
 patience, research and, 70

 Pauli, Wolfgang, 288, 335
 Pauli exclusion principle, 335, 362
 Payne-Gaposchkin, Cecilia, 301
 peculiar galaxies, 430–431
 Peebles, P.J.E., 446
 Pegasus constellation, 351
 quasar, 438
 penumbra, 19, 20, 21
 sunspots, 274
 penumbral eclipse, 19
 Penzias, Arno, 446
 peridot, 249
 perihelion, 38
 periodic table of elements, 85, 86
 period–luminosity relation, 337–338
 for Type I Cepheids, 391–392
 Permian-Triassic boundary, 263
 peroxides, 183
 Perseids meteor shower, 258
 Perseus Arm, 396
 Perseus cluster of galaxies, 432, 436
 Perseus constellation, 343
 Petit-Prince (moon), 242
 phase transition, 196
 phi (ϕ) Persei, 344
 Phobos (moon), 183–184
 Phoebe (moon), 215, 216
 Phoebe ring, 213
Phoenix Mars Lander, 171, 175
 phosphorus, and existence of life, 471
 photodisintegration, 356
 photoelectric effect, 54
 photography, and astronomy, 66–68, 239
 photoionization, 90, 92
 photometry, 300
 photomultipliers, 289
 photon collisions, 453
 photon energy, 54
 photon pressure, in comets, 251
 photons, 50, 54
 absorbed and emitted, 90–92
 early universe, 452–453
 formation of, 451
 frequency of, 54
 ionization, 454
 long- and short-wavelength, 83
 primordial, 455
 photosphere (Sun), 268, 269, 270, 277, 278, 279, 283
 photosynthesis, 130
 physical forces, unification of, 449, 450
 Piazzi, Giuseppe, 237
 Pickering, Edward C., 302
 Pierre Auger Observatory, 362
 Pila Patera (Io), 201
Pioneer 10, 478, 479
Pioneer 11, 478, 479
Pioneer spacecraft, 198
Pioneer Venus Orbiter, 162
 Pisces (the Fish) constellation
 Sun in, 9, 10
 vernal equinox in, 16
 Pistol Star, 326
 pixels, 67–68
 PKS 2000-330 (quasar), 429
 plagues, 279–280
 Planck, Max, 54, 450
 ideal blackbody curves, 83

- Planck era, 449, 450, 465
- Planck's constant, 54
- Planck's law, 83
- Planck Space Telescope, 457
- Planck time, 450
- planetary configurations, 33–35
- planetary differentiation, 135
- planetary habitable zone, 475
- planetary motion
 - geocentric explanation, 32–33
 - heliocentric explanation of, 31–36
- planetary nebulae, 99, 349–351
 - shapes of, 351
- planetary orbits, 108, 109
- planetesimals, 238
 - in Kuiper belt, 105
- solar system formation, 103, 104
- planetology, comparative, 126–129, 184–185
- planets, 30, 97, 234
 - albedos of, 128
 - average density of, 127, 128
 - configurations of, 33–35
 - conjunction and opposition of, 35
 - defined, 108
 - direct and retrograde motion of, 30–33
 - elongations of, 33–35
 - formation of, 103–105
 - free-floating, 120
 - inferior and superior, 33
 - mass of, 126–127
 - moons of, 128–129
 - number of, 122
 - orbit of, 37–39, 126
 - periods of, 35, 40–41
 - proper motion of, 92
 - sidereal and synodic periods of, 18, 35, 39
 - sizes of, 126, 127
 - spectra of, 128
 - speed of, 38, 39
 - see also* dwarf planets; exoplanets
- plasmas, 454
 - Jupiter, 207
 - photosphere, 277
- plasma tail (of a comet), 251–252
- plasma tori, on Jupiter, 207
- plate tectonics, 133–135, 138
 - on Ganymede, 204
- Pleiades star cluster, 311, 318, 320, 322, 329, 339
- plutinos, 105, 106, 246, 247
- Pluto, 234–237, 240
 - and Charon, 235–237
 - demotion from planet status, 108
 - formation of, 105, 106
 - mass of, 235
 - and Neptune's gravity, 247
 - orbital eccentricity, 234–235
- plutoids, 105, 106, 234
- pointer stars, Big Dipper as, 3, 7
- Polaris (star), 3–4, 8, 14
- pollution, light, 2
- polycyclic aromatic hydrocarbons (PAHs), 473
- polymers, 218
- poor cluster, 414
- Population I stars, 340, 342, 351, 396
- Population II stars, 340, 396
- Population III stars, 456–457
- positrons, 285, 399, 453
 - early universe, 452
 - from proton-proton fusion, 284, 285
- powers-of-ten notation, 23
- precession, 14–16
 - of equinoxes, 16
- predictions, 28–30
 - quantitative, 46
- pre-main-sequence stars, 324–326
- pre-solar nebula, *see* solar nebula
- primary cosmic rays, 361
- primary mirrors, 58, 59, 60, 62, 63, 67, 70–71
 - light-gathering power of telescope and, 60, 61
- prime focus, 59
- prime meridian, 7
- primordial atom of universe, 446
- primordial black holes, 378, 379–380, 383
- primordial fireball, 454
- primordial gas, galaxies formed from clouds of, 455–459
- primordial nucleosynthesis, 453–454
- primordial photons, 447, 455
- prism, spectrum and, 50, 51
- Procyon (star), 4
- prograde motion, 30
- prograde orbits, 206
- Project Ozma, 471, 476
- Prometheus (moon), 201, 212, 213
- prominences (of filaments), 279, 280
- proper motion, 92–93
- proplyds, *see* protoplanetary disks
- protogalaxies, 458, 459
- proton-proton chain (PP chain), 284–285, 333
- protons, 89–90
 - early universe, 452, 453
- protons and electrons gas, and black holes, 435–437
- protoplanetary disks, 102, 114
 - distorted, 114–115
- proto-solar nebula, *see* solar nebula
- protostars, 323–324, 325
 - in Bok globule, 323
- protosun, 100, 101, 102–104, 110
- Proxima Centauri, 40, 294, 296, 299
- PSR 0329+54 (pulsar), 362
- PSR 1916+13 (pulsar), 367
- PSR 1937+21 (pulsar), 364
- PSR B1257+12, 118
- PSR J0337+1715 (millisecond pulsar), 367
- PSR J0737-3039 (double pulsar), 367
- Ptolemy, 32
- pulsars, 118, 220–222, 362–369
 - glitch in, 366–367
 - lighthouse model of, 365
- pulsar timing method for detecting planets, 117
- pyroxene, on Mars, 180
- Pythagoras, 28
- Quadrantids meteor shower, 258
- quantum fluctuations, 456
- quantum mechanics, 88, 90
- Quaoar, 247
- quarks, 449, 450, 451
 - early universe, 452
- quasars, 426–430
 - brightness of, 429–430
 - central engine, 437
 - earliest, 458
 - energy emitted by, 429
 - formation of, 429
 - gravity and light from, 437–438
 - luminosity of, 429–430, 432
 - redshifts of, 427–430
- quasi-stellar objects (QSOs), 428
- quasi-stellar radio sources, *see* quasars
- quintessence, 465
- Quintuplet Cluster, 326
- R136 (star cluster), 317
- R136a1 (star), 326
- radial velocity, 92, 93
- radial-velocity curve, 311
- radial velocity method of detecting exoplanets, 115–116
- radiation-dominated universe, 454
- radiation pressure, in comets, 251
- radiative transport, 283
- radiative zone (Sun), 283, 286
- radioactive age-dating, 89–90
- radioactive decay, 89, 135
- radioactive elements, 88
- radioactive isotopes, 89
- radio galaxies, 430–432
- radio intensity, 74
- radio lobes, 427, 431
- radio radiation, 431
- radio sources, 426–428
- radio stations, frequency of, 55
- radio telescopes, 72–75
 - mapping Milky Way, 394, 395
- radio waves, 55, 56, 72–73, 80
 - in early universe, 446
 - and search for extraterrestrial intelligence, 471, 474–477, 479
- raindrops, on Titan, 216
- RCW 49 (star), 324
- Reber, Grote, 426
- recessional velocity, 422
- recombination, 350
 - era of, 455
- red dwarfs, 304, 332–333, 348
- red giant stars, 304, 334
- Red Sea, 134
- redshift, 54
 - and clusters of galaxies, 445
 - cosmological, 445, 446
 - of quasars, 427–430
 - of superclusters, 422–425
- Red Spot Jr., 194, 195
- red supergiants, 350
- reflecting telescopes, 57–60, 63
 - disadvantages of, 64–66
 - infrared and ultraviolet, 75–77
- reflection, principle of, 57–58
- reflection nebula, 318, 321
 - Pleiades, 320
- reflectors, *see* reflecting telescopes
- refracting telescopes, 57, 63–64
 - essentials of, 64
 - largest, at Yerkes Observatory, 63, 65
 - problems of, 63–64

- refraction, 50, 62–63
- regolith, 259
- regular cluster of galaxies, 415
- Regulus (star), 4, 300, 306
- relativity, 370
- relativity theories, 282, 369–374
 - see also* general relativity, theory of; special relativity
- remote binaries, 311
- Remus (moon), 242
- repulsive gravity, 464
- resolution, *see* angular resolution
- resonance, 105
 - Cassini division, 211–212
 - plutinos and Neptune, 247
- retrograde motion, of planets, 30–33
- retrograde rotation, 168
 - Uranus, 220
 - Venus, 168
- revolution, 9–10
- Rhea (moon), 213, 218
- rich cluster, 414
- right ascension (R.A.), 6
- rilles, 143–144
- ringlets
 - Jupiter, 206
 - Saturn, 211, 212
- Roche, Edouard, 340
- Roche limit, 213
- Roche lobes, 342–343, 368
- rocket nozzle model, 436
- Roentgen, Wilhelm, 55
- Roman Catholic Church, Galileo's
 - condemnation by, 41
- Rømer, Ole, 53
- Romulus (moon), 242
- Rosetta* spacecraft, 252, 253
- Rosette Nebula (NGC 2237), 67, 92, 323
- Rosse, Lord, *see* Parsons, William, Lord Rosse
- rotating radio transients (RRATs), 366
- rotation, in solar system formation, 102
 - see also* differential rotation; retrograde rotation
- RR Lyrae stars, 392
- RR Lyrae variables, 337
- rubidium, 85
- Russell, Henry Norris, 303–304
- Rutherford, Ernest, 88, 89

- Sa galaxies, 404, 407
- Sagan, Carl, 470
- Sagitta (the Arrow), 4
- Sagittarius A (Sgr A), 398, 426
- Sagittarius A, 398, 399
- Sagittarius Arm, 396
- Sagittarius constellation
 - globular clusters, 393
 - radio noise, 72
- Sagittarius Dwarf galaxy, 400, 414–415
- Saha, Meghnad, 301
- Sakharov, Andrei, 456
- salts, on Mars, 177
- San Andreas fault, 134, 136
- Sandage, Allan, 427
- sand dunes, on Titan, 218
- Santa Maria Rupes (Mercury), 158
- Sapas Mons (Venus), 166

- Saturn, 191, 207–220
 - albedo of, 128
 - atmosphere, 207–210, 228
 - auroras of, 207, 208
 - average density of, 128
 - belts and zones on, 207–208, 209
 - differential rotation of, 207–208
 - formation of, 104, 105
 - interior of, 196, 210, 228, 229
 - magnetic field of, 215, 223, 228, 229
 - mass of, 127, 228
 - moons of, 129, 215–220
 - orbital data, 126
 - rings of, 210–215, 220
 - rotation of, 228
 - size of, 126, 127, 228
 - storms and winds on, 208–209
 - synodic and sidereal periods, 35
 - temperatures on, 195
 - visible and radio views of, 74
 - vital statistics, 208
- Saturn–Sun distance, 35, 126
- SBa galaxy, 409
- SBb galaxy, 409
- SBc galaxy, 409
- Sb galaxies, 404, 407
- scorpius
 - on Mercury, 158, 159
 - on Moon, 148
- scattered KBOs, 247
- Sc galaxies, 404, 407
- Schaller, Adolf, 443
- Scheila (asteroid), 241
- Schiaparelli, Giovanni Virginio, 169
- Schmidt, Bernhard, 65
- Schmidt, Maarten, 428
- Schmidt–Cassegrain telescopes, 66
- Schmidt corrector plate, 65–66
- Schmitt, Harrison, 146
- Schwabe, Samuel, 273
- Schwarzschild, Karl, 374
- Schwarzschild black holes, 375, 435
- Schwarzschild radius, 374, 375, 376
- science, 28–30
 - as inclusive, 30
- scientific method, 28–30
- scientific notation, 23
- scientific theory, 29–30, 134
 - explanations and, 41
- scintillators, 77
- Scorpius–Centaurus OB association, 359
- Scott, David, 144
- Scutum–Centaurus Arm, 396
- seafloor spreading, 134
- Search for Extraterrestrial Intelligence (SETI), 123, 471, 474–477
- seasons
 - on Earth, 10–16
 - on Uranus, 220, 222
- secondary cosmic rays, 361
- secondary mirror, 58, 59–60
- Sedna (dwarf planet), 247, 248
- Sedna (Oort cloud object), 106
- seeing, 68
- seeing disk, 68
- Segue 1, 420
- seismic waves, 137
- seismographs, 137

- SELENE* orbiter, 145
- self-propagating star formation, 406
- semidetached binary systems, 342, 343, 344
 - white dwarfs, 353
- semimajor axis, 37, 38, 39
- SETI@home, 476
- seven sisters (Pleiades), 311
- Seyfert, Carl, 430
- Seyfert galaxies, 430, 432, 436
- Shapley, Harlow, 391, 393
- Shapley concentration, 448
- Shapley–Curtis debate, 391–392
- shepherd satellites
 - Saturn, 212, 213
 - Uranus, 222–224
- shield volcanoes, 174
- shock waves
 - galaxy collisions, 416
 - OB associations, 328
 - and star birth, 406
 - supernovae, 356
- Shoemaker, Carolyn and Eugene, 198, 246
- shooting stars, 256–257
- short-period comets, 254
- Shu, Frank, 407
- Siberia, Tunguska event, 257
- sidereal day, 13, 14
- sidereal month, 17–18
- sidereal period, 9, 35, 39
- silicon fusion, 355, 356
- silicon-oxygen bonding, 471, 472
- silicon-silicon bonding, 471
- single-line spectroscopic binary, 311
- singularity, 373, 375, 377
- Sirius (star), 4, 297, 352
- Sirius B (star), 352
- Sitter, Willem de, 461
- 61 Cygni (star), 294
- Skylab* missions, 77
- Slipher, Vesto, 422
- Sloan Digital Sky Survey, 413, 438
- Small Magellanic Cloud, 391, 411
- small solar system bodies (SSSBs), 108, 234, 238–264
 - see also* asteroids; comets; meteorites; meteors; meteoroids
- Smoot, George F., 447
- SN 1997ff, 463
- SNC meteorites, 259
 - from Mars, 181, 182
- snow line, 103, 104
- sodium vapor lights, 87
- SOFIA (Stratospheric Observatory for Infrared Astronomy), 76
- soft gamma-ray repeaters (SGRs), 365
- Sojourner* rover, 179
- Solar and Heliospheric Observatory (SOHO), 77, 255, 280, 281, 282
- solar corona, 20, 21, 271–272, 279, 280
- solar cycle, 278–279
- solar day, 13, 14
- solar disk, 269
- Solar Dynamics Observatory (SDO), 77, 267
- solar eclipses, 18, 20–22
 - in 2015–2018, 22
 - and chromosphere, 270
- solar flares, 281
- solar granulation, 269, 270, 274

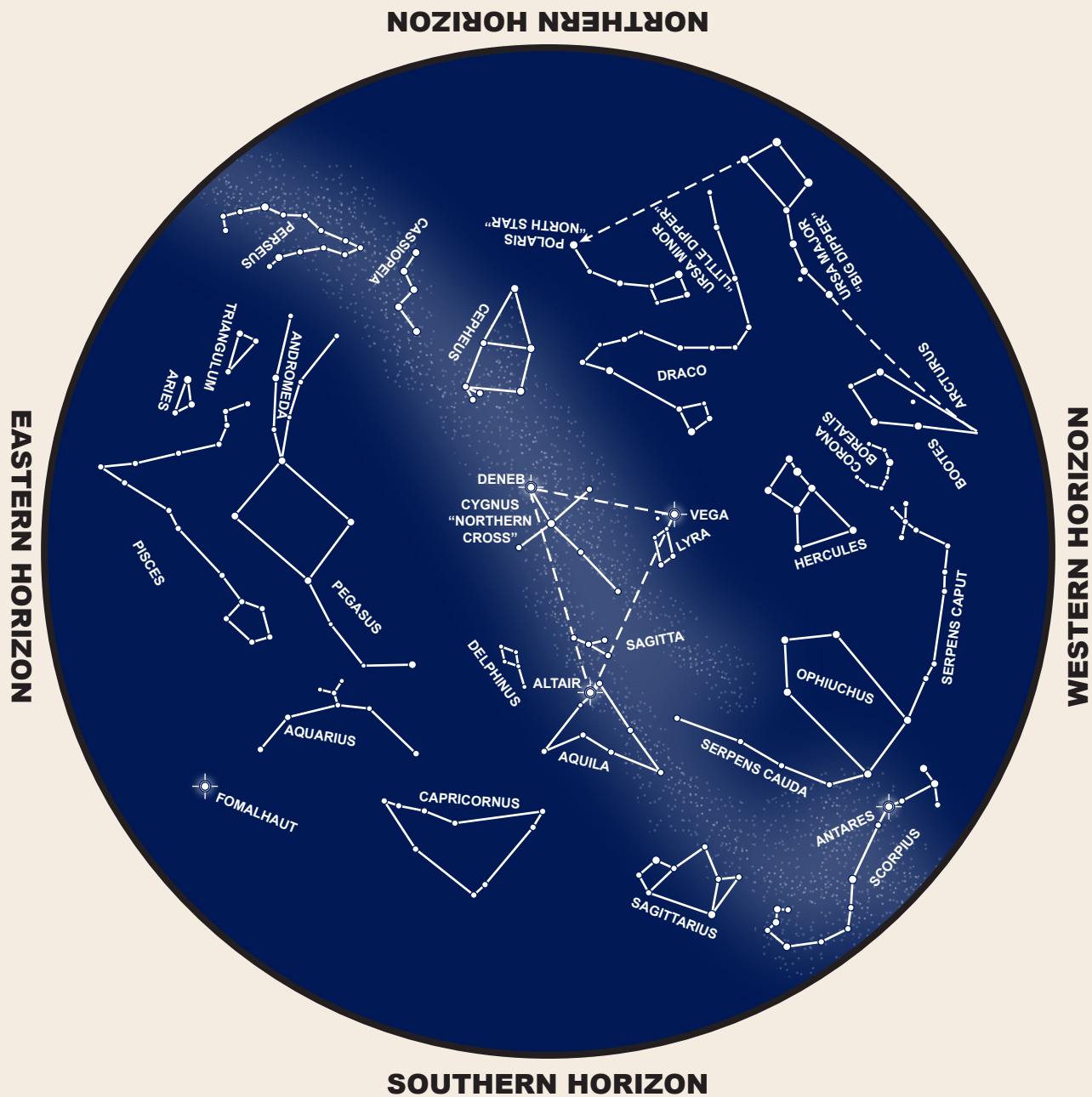
- solar luminosity, 283, 284
- solar maximum, 279
- solar nebula, 100–102
- solar neutrinos, 287–289
- solar radius, 286
- solar spectrum, 84, 85, 87–88, 276, 342
- solar system
 - age of, from meteorites, 259
 - birth of, from supernova, 262
 - categories of objects, 108
 - comet reservoirs, 246–248
 - formation of, 98–103, 236
- solar telescopes, 77
- Solar Terrestrial Relations Observatory (STEREO), 77
- solar vibrations, 277–278
- solar wind, 139–140, 250, 251, 272–273, 280
 - and Jupiter, 197
 - and Mercury, 159
- solstices, 11
 - see also* summer solstice; winter solstice
- Sombrero Galaxy (M104), 389, 404, 434
- sound, speed of, 53
- sound waves, in inflationary universe, 456
- south celestial pole, 5, 6, 14
- southern lights (aurora australis), 140
- Soviet Union
 - Moon missions, 145
 - Venus probes, 163
- spacetime, 371
 - curvature of, 371, 372
 - expansion of, 425, 445
 - formation of, 446
 - and gravitational waves, 374
 - and light, 372–373
 - and matter, 371
 - shape of, 461–463
- space weather, 282
- space weather satellites, 140–141
- special relativity, 369–371
- spectra, 50, 51
 - of planets, 128
- spectral analysis, 84–88
 - of binary stars, 311
- spectral lines, 84–87, 276–277
 - brightness levels of, 87–88
 - orbital motion of binary stars and wavelengths of, 311–312
 - Zeeman effect, 276, 277
- spectral types, 302–303
- spectrogram, 84
- spectrograph, 59, 85–86
 - grating, 86
- spectroscope, 84
- spectroscopic binaries, 311–312
- spectroscopic parallax, 306, 424
- spectroscopy, 84, 85, 311
 - stellar, 301
- speculum metal, 59
- spherical aberration, 64, 65, 66
- spherical lens, geometry of, 65
- Spica (star), 4
- spicules, 270, 271
- spin, 394
- spin flips, 394
- spiral arms, 406–407
 - barred spiral galaxies, 395, 396, 409
 - of flocculent spirals, 406–407
 - Milky Way, 395, 396
 - spiral galaxies, 403–405
- spiral density waves, 406
- spiral galaxies, 403–405, 411, 412
 - early universe, 459–460
 - formation of, 412
 - rotation curves of, 420
 - see also* Seyfert galaxies; barred spiral galaxies; flocculent spiral galaxies
- spiral nebulae, 391
- spiral tracers, 395
- Spirit* rover, 170, 176, 177, 183
- Spitzer Space Telescope, 75–76, 213
 - and ingredients for life, 473
 - Milky Way, 390, 396
 - quasars, 437
- spokes, in Saturn's rings, 215
- spring, 11
- spring tides, 153
- stable Lagrange points, 244
- standard candles, 423
 - Type Ia supernovae, 463
- star birth, 322, 416
- starburst galaxies, 416, 417
- star clusters, plotting on H-R diagrams, 328–330
- Stardust* mission, 248–249, 472
- star formation, 395
 - early universe, 443, 456–458, 459–460, 464
 - in flocculent spirals, 406
 - and galactic collisions, 416
 - in merging galaxies, 419
 - process summary, 331
 - self-propagating, 406
- Starry Night*TM, 6
- stars
 - age of, 328–330
 - apparent magnitude of, 297–298
 - Balmer lines of, 302–303
 - blackbody spectra of, 300
 - circumpolar, 8
 - classification of, by spectra, 302–303
 - color of, and temperature, 300–301
 - first, and solar system formation, 98–99
 - formation of, 98–99
 - fully convective, 332
 - giant phase, 333–334
 - lives of, 317–318
 - lowest mass, 332–333
 - of more than 0.4 M, 333–334
 - motion of, at poles, 8
 - with multiple planets, 120
 - neutron, 118
 - open clusters of, 99
 - proper motion of, 92–93
 - radial and transverse velocity of, 93
 - red dwarfs, 348
 - red supergiants to white dwarfs, 350
 - rising and setting of, 8, 9
 - temperatures, 299–303
 - types of, 303–306
 - variable, 336–337
 - see also* entries under stellar and binary star systems; binary stars; neutron stars; supergiant stars
- stars, high-mass, 349, 354–362
 - evolutionary stages, 356
 - neutron degeneracy pressure, 373
 - structure of, 355
 - and supernovae, 355–358
- stars, intermediate-mass, 348–354
 - evolution of, 348
 - structure of, 349
- steady-state theory, 447
- Stefan, Josef, 82
- Stefan-Boltzmann law, 82, 269, 304
- stellar black holes, 383
 - in Milky Way, 377–378
- stellar cataclysms, 382
- stellar cores, and white dwarfs, 351–352
- stellar distances, 294–296
- stellar evolution, 294
 - summary, 384
- stellar masses, 307–312
- stellar nurseries, 322–323
- stellar parallax, 294–296
- stellar radii, 305
- stellar remnant black holes, 377–378
- stellar remnants, 362, 364
- stellar spectra, 305
 - and gravitational redshifts, 373
 - and luminosity, 305–306
 - of metal-rich and metal-poor stars, 342
 - and star surface temperature, 301–302
- stellar spectroscopy, 301
- stellar structure, equations of, 287, 325
- stellar winds, 99
 - giant stars, 349, 350
 - intermediate-mass stars, 349
 - OB associations, 328
- Stephan, Edouard, 412
- Stephan's quintet, 412
- Stickney, Angeline, 184
- Stickney crater on Phobos, 184
- stones, 260
- stony-iron meteorites, 261
- stony-irons, 260
- stony meteorites, 260–262
- storms, on Saturn, 208–209
- stratosphere, 131, 132
- strong nuclear force, 89
 - and other physical forces, 449, 450, 451
- Styx (moon), 236
- Subaru Telescope, 75
- sublimation., 161
- Sudbury Neutrino Observatory, 288–289
- sulfur, 201, 202
- sulfur dioxide
 - on Io, 202
 - on Venus, 163
- sulfuric acid, on Venus, 163
- summer, 11
- summer solstice, 11, 12
- summer triangle, 4–5
- Sun, 24, 267–291
 - absolute magnitude of, 298
 - absorption spectrum of, 85, 87–88
 - active, 273–282
 - apparent magnitude of, 297
 - atmosphere of, 269–273
 - as a blackbody, 80, 82
 - core of, 282–283, 286, 287
 - daily path and energy deposited, 11–12
 - development of, 110
 - differential rotation of, 275, 277, 278, 401
 - ecliptic, 6–7, 9, 10

- energy of, 282–287
- evolution of, 336
- fate of, 351
- formation of, 102
- as G2 star, 303, 304, 306
- as a giant, 334
- gravitational pull on Moon, 18
- gravity of, 272
- increasing brightness of, 287
- interior of, 282–289
- iron in, 85
- luminosity of, 432
- magnetic fields of, 272–273, 276–282
- mass of, 287
- and Moon, 16
- orbital speed, 401
- as Population I star, 340
- rotation of, 277, 364
- temperatures of, 271–273, 279, 287
- ultraviolet image of, 267
- vital statistics, 268
- see also entries under solar*
- sun-centered cosmology, *see* heliocentric cosmology
- sungrazing comets, 255, 256
- sunquakes, 277
- sunspot collisions, 281
- sunspot cycle, 273–276, 278, 279
 - and solar flares, 281
- sunspot groups, 273, 274, 275
 - and solar flares, 281
- sunspot maximum, 274, 275
- sunspot minimum, 274, 275
- sunspots, 273–279
 - granulation, 274
 - magnetic fields of, 276
 - magnetic north and south poles, 278
 - and Sun's magnetic fields, 276–279
- superclusters (of galaxies), 24, 456, 457
 - distribution of, 412–413
 - gravitational binding of, 422
 - motion of, 422–426
 - redshifts of, 422–425
 - separation rate, 464
- superconductors, 366
- supercontinents, 137
 - Pangaea, 134, 135
- super-Earth, 119
- superfluids, 366
- supergiant stars, 304, 305, 349–351, 355, 360
- supergranules, 270, 271
- superior conjunction, 33
- superior planets, 33
- supermassive black holes, 362, 378–379, 381, 399, 401, 433–435
 - accretion disks, 437
 - formation of, 457–458, 460
 - and galactic activity, 435, 436
 - galaxy structures and, 412
 - gravitational energy, 433–435
 - head-tail sources, 432
 - merger of, 419
- supermassive engines, 433–438
- supermassive stars, mass loss from, 326
- Supernova 1987A (SN 1987A), 359–360, 361
- supernovae, 99, 322–324
 - and gamma-ray bursts, 382
 - of high-mass stars, 355–358, 373
 - light curve of, 354
 - shock waves from, 406
 - supernova explosion, 262
 - supernova of 1572, 36, 37, 347, 359
 - supernova remnants, 322, 358–362, 398, 426
 - superoxides, 183
 - superstar cluster, 340
 - superstring theory, 373–374, 449
 - SWIFT Gamma-Ray Burst Mission, 77, 382
 - symmetry breaking, early universe, 453
 - synchronous rotation, 474
 - of Moon, 152–153, 154
 - synchrotron radiation, 398, 431
 - synodic month, 17–18
 - of a planet, 35
 - Syrts Major (Mars), 171
 - Tarantula Nebula, 317, 360, 411
 - Tau Bootis (star), 278
 - Tau (τ) Ceti (star), 471–472
 - tau neutrinos, 288
 - tau particles, 288
 - Taurids meteor shower, 258
 - Taurus (the Bull), 4, 320
 - see also* Pleiades star cluster
 - Taurus-Littow Valley (Moon), 146
 - Taylor, Joseph, 374
 - tectonic plate motion
 - Enceladus, 219
 - on Mars, 174
 - tectonic plates, 134–135, 136, 137
 - telescope mirrors, 64–66
 - telescopes
 - functions of, 60–62
 - light-gathering power of, 61
 - and nonvisible light, 71–79
 - see also* reflecting telescopes; refracting telescopes
 - temperature, and color and brightness, 79–83
 - terminator (moon), 17
 - terraces (in craters), 142
 - terrestrial planets, 105
 - testing, 29
 - Tethys (moon), 216
 - Tharsis Bulge (Mars), 170
 - Thebe (moon), 207
 - theories, 29
 - beliefs and, 30
 - explanations and, 41
 - Theory of Everything, 449
 - thermal pulses, 352
 - thermal runaway, 350
 - thermonuclear fusion, 98–99, 283, 284–285
 - thermophiles, 476
 - thermosphere, *see* ionosphere
 - Third Cambridge Catalogue*, 427
 - third quarter Moon, 16
 - 3C 48 (quasar), 427
 - 3C 75, 432
 - 3C 273, 428
 - 3C 279, 429
 - 3C 294 (galaxy), 458
 - 3-degree background radiation, 447
 - three-star systems, 367
 - 3-to-2 spin orbit coupling, 160, 161
 - Thuban, 14
 - thulium, 367
 - tides, 152, 153–154
 - on Uranus, 223
 - tiger stripes, 219–220
 - time
 - and black holes, 376
 - Earth's rotation and revolution, 13–14
 - slowing down of, 372
 - time dilation, 370–371, 377
 - Tissint meteorite, 259
 - Titan (moon), 192, 213, 214, 215–218
 - size of, 126
 - water in, 110
 - Toby Jug (star), 334
 - Tombaugh, Clyde, 234
 - total eclipse, 19–20
 - totality (of solar eclipse), 21
 - total solar eclipses, 20–21
 - in 2001–2020, 22
 - Toutatis (asteroid), 241, 242
 - trailing-arm spiral galaxies, 409
 - transform-fault boundaries, 134
 - transitions, and electron orbits, 90
 - transition zones, chromosphere and corona, 271, 272
 - transit photometry method (to detect planets), 115, 116, 119
 - trans-Neptunian objects (TNOs), 106, 108, 237
 - transverse velocity, 92, 93
 - Trapezium, 113, 319, 328
 - Triangulum galaxy, 381
 - triple alpha process, 335–336, 350
 - A Trip to the Moon* (movie), 141
 - Triton (moon), 228
 - Trojan asteroids, 239, 241
 - tropical year, 14
 - troposphere, 131
 - true vacuum state (of universe), 451
 - Trumpler, R.J., 392–393
 - T Tauri stars, 325, 326
 - tube worms, 476
 - Tully, Brent, 424
 - Tully–Fisher relation, 424
 - Tunguska event, 257
 - turbulence, 68
 - on Jupiter, 193, 194
 - turnoff point, 339
 - Tvashtar Catena (Io), 201
 - 21-cm emission line, of spiral galaxy, 424
 - 21-cm radio radiation, 394, 395
 - twinkling, 68, 69
 - Two Degree Field (2dF), Survey, 413
 - 2M1207, 116, 117
 - 2M1207b, 116, 117
 - Type I Cepheid variables, 337–338
 - Type I Seyfert galaxies, 430
 - Type Ia supernovae, 353–354, 358
 - dark energy, 463, 465
 - and distance measuring, 424
 - Type II Cepheid variables, 337–338
 - Type II supernova, 354–364
 - Type 2 Seyfert galaxies, 430
 - Uhuru* satellite, 367–368
 - ultracompact dwarf galaxies, 410
 - ultrahigh-energy cosmic rays, 362
 - ultraviolet photons, 183

- ultraviolet (UV) radiation, 55, 56, 57, 130
 - and ozone layer, 131
- ultraviolet telescopes, 75, 77
- Ulysses* spacecraft, 281
- umbra, 19, 20, 21
 - sunspots, 274
- unidentified flying objects (UFOs), 470, 471
 - Venus as, 162–163
- universe
 - acceleration outward, 444, 463–465
 - age of, 445–446
 - closed, 462
 - defined, 2, 444
 - early evolution of, 446, 450–455, 456, 460
 - evolution of, 450, 464
 - expansion of, 422–425, 444–446, 450, 456
 - expansion of, and general relativity, 461
 - fate of, 460–465
 - and galaxies, 391
 - geocentric view of, 30–31
 - geometries of, 461
 - gravitational attraction in, 454
 - and life, 122, 123
 - major components of, 465
 - matter-dominated, 454
 - matter formation in, 453
 - observable, 450, 451, 452
 - open, 462
 - radiation-dominated, 454
 - scales of, 23
 - shape of and fate of, 461–463
 - sound waves in, 456
 - structure of, 455–460
- Upsilon Andromedae, 121
- uranium, 85, 89
- Uranus, 220–224
 - albedo of, 128
 - atmosphere of, 220, 229
 - average density of, 128
 - belts and zones of, 220
 - clouds, 220, 222
 - discovery of, 45
 - formation of, 104
 - gravity and comets, 246
 - interior of, 220–222, 228, 229
 - magnetic field of, 220–222, 223, 228, 229
 - mass of, 127, 228
 - moons of, 129, 220, 222–224
 - orbital data, 126
 - rings of, 222–224
 - rotation of, 220–222, 228
 - size of, 126, 127, 228
 - synodic and sidereal periods, 35
 - vital statistics, 221
 - water on, 220
- Uranus–Sun distance, 35, 126
- Urey, Harold, 473
- Ursa Major constellation
 - black hole, 380
 - spectrum of, 422
- Ursids meteor shower, 258
- UVA, UVB, and UVC radiation, 57
- V404 Cygni (black hole), 378
- V461 Sgr, 378
- V838 Monocerotis (red giant), 293
- vacuum energy, 464–465
- Valhalla Basin (Callisto), 205
- Valhalla Basin, 206
- Valles Marineris (Mars), 169–171, 179
- Van Allen, James, 140
- Van Allen radiation belts, 140
 - solar wind, 281
- variable stars, 336–337
 - see also* Cepheid variable stars
- Vastitas Boreali (Mars), 169, 171, 182
- Vega (star), 4, 14, 91, 114, 297
- Vela constellation, 358
- Vela pulsar, 365, 366
- Vela satellites, 381
- velocity, 43
- Venera* 7, 163, 164, 167
- Venera* 13, 165
- Venus, 162–168
 - air pressure on, 162, 163
 - albedo of, 128
 - apparent magnitude of, 297
 - atmosphere of, 163, 164–165, 185
 - cloud cover of, 162–165
 - crust of, 167
 - elongation of, 33, 35
 - formation of, 104–105
 - impact craters on, 166–167
 - interior of, 167–168, 185
 - lava on, 165, 166
 - magnetic fields, 163, 185
 - mantle of, 167
 - map of, 166
 - mass of, 127, 185
 - maximum elongation of, 38
 - orbital data, 12
 - phases of, 41
 - retrograde rotation of, 168, 185
 - rotation axis, 168
 - size of, 127, 185
 - sulfur on, 163
 - surface of, 162–168, 185
 - synodic and sidereal periods of, 35
 - temperatures on, 163, 164–165
 - vital statistics, 162
 - volcanoes on, 163, 165–168
 - water on, 165
- Venus Express*, 167
- Venus–Sun distance, 35, 126
- VERITAS array, 78
- vernal equinox, 6, 7, 11, 12, 16
- Verne, Jules, 141
- Very Large Array (VLA), 74, 398, 417
 - Cassiopeia A, 359
 - Cygnus A, 427
- Very Large Telescope, 49, 71
- Very Long Baseline Array (VLBA), 74
- very-long-baseline interferometry (VLBI), 74
- Vesta (asteroid), 107, 239, 240, 256
 - meteorites from, 259
- Viking 1* and *Viking 2 Landers*, 170, 183
- Viking 1* and *Viking 2* orbiters, 171, 172, 178, 182
- Viking* Mars Lander, 470
- Virgo (the Virgin), 4
 - galaxies, 409, 410, 413
 - spectrum of, 422
 - Sun in, 9, 10
- Virgo cluster of galaxies, 414, 415, 435, 448
- Virgo Supercluster, 413
- virtual particles, 383, 456
- visible light, 50, 53, 55, 56, 57, 80
 - from solar corona, 272
 - wavelengths of, 52
- visual binaries, 307, 311
 - center of mass of, 308–309
- voids, 456, 460
 - and superclusters, 413–414
- volcanism, water, 203
- volcanoes, 173
 - at Earth's boundaries, 134
 - on Io, 201–202
 - on Mars, 169, 171, 172, 173–174
 - on Mercury, 156, 157, 158
 - on Venus, 163, 165–168
- Voyager 1* and 2 missions, 192, 193, 478, 479
 - Callisto, 205
 - Galilean moons, 200, 201, 202
 - Miranda, 225
 - Neptune, 226
 - Saturn's rings, 212, 214
 - and Sun, 273
 - Titan, 216, 217
 - Triton, 227, 228
 - Uranus, 220, 223
- Vulpecula constellation, black hole, 379
- Walker, Arville, 302
- waning crescent Moon, 16
- waning gibbous Moon, 16
- waning Moon, 17
- War of the Worlds* (Wells), 169
- WASP-12b, 120
- water
 - on Callisto, 205
 - on Enceladus, 219–220
 - on Europa, 202–204
 - on Ganymede, 204
 - on Mar, 172, 174, 175, 177–182
 - on Mars, 169, 171
 - in meteorites, 261
 - on Titan, 218
 - on Uranus, 220
 - on Venus, 165
- water, liquid
 - Earth, 130, 287
 - on moons, 474
 - planets with, discovery of, 122
 - solar system formation, 105, 110
- water-bearing rock, on Moon, 148
- water hole, 476
- water ice
 - on Mars, 181
 - on Moon, 148–150
- water vapor
 - on Earth, 130
 - and infrared radiation, 75–76
 - on Jupiter, 195
- water volcanism, on Europa, 203
- wavelengths, of colors, 51, 52, 53
- wave-particle duality, 54
- waxing crescent Moon, 16
- waxing gibbous Moon, 16
- waxing Moon, 17
- weak nuclear force, 89
 - and neutrons, 288
 - and other physical forces, 449, 450, 451

- Wegener, Alfred, 134
 weight, 41
 Wells, H. G., 169
 West, Michael M., 250
 Westerlund 1 (star cluster), 339–340
What If the Earth Had Two Moons?, 152
 white dwarfs, 304, 305, 351–354, 364
 in close binary systems, 353–354
 and dark matter, 402
 light from, 373
 white light, 50, 51
 white ovals, 193, 195
 Whyte, Marion, 302
 Widmanstätten, Count Alois von, 260–261
 Widmanstätten patterns, 261
 Wien, Wilhelm, 82
 Wien's law, 82, 83
 cosmic background radiation, 447, 455
 Wilkinson, David, 446
Wilkinson Microwave Anisotropy Probe
 (WMAP), 74, 446, 447, 456
 William of Occam, 29
 Wilson, Robert, 446
 winding dilemma, 405
 windows, in Earth's atmosphere, 57
 winter, 11
 winter solstice, 11, 12
 winter triangle, 4
 wobbling (of stars), and discovery of
 exoplanets, 116
 Wolf-Rayet stars, 382
 Woods, Ida, 302
 wormholes, 377
 W Ursae Majoris, 343, 344
 Xanadu highlands (Titan), 217
 XMM-Newton Telescope, 77
 X-ray burster, 369
 X-ray nucleus, 431
 X-ray photons, 77
 X rays, 55, 56, 77, 80
 from black holes, 398–399
 and dark matter, 420
 X-ray sources, 77, 380
 M87, 434
 pulses from, 367–369
 X-ray telescopes, 77, 78
 grazing incidence, 77, 78, 79
 year, 9, 13–14
 Yerkes Observatory, 63, 65
Yohkoh satellite, 272
 Young, Thomas, 50–51, 52
 Yucatán Peninsula asteroid,
 263, 264
 Zeeman, Pieter, 277
 Zeeman effect, 276, 277
 zenith, 6
 zero-age main sequence (ZAMS), 330
 Zeta Leporis (star), 114, 244
 zeta (ζ) Orionis, *see* Alnitak
 zodiac constellations, 13
 zonal flow, 193
 zones, *see* belts and zones
 Zwicky, Fritz, 362

This page intentionally left blank



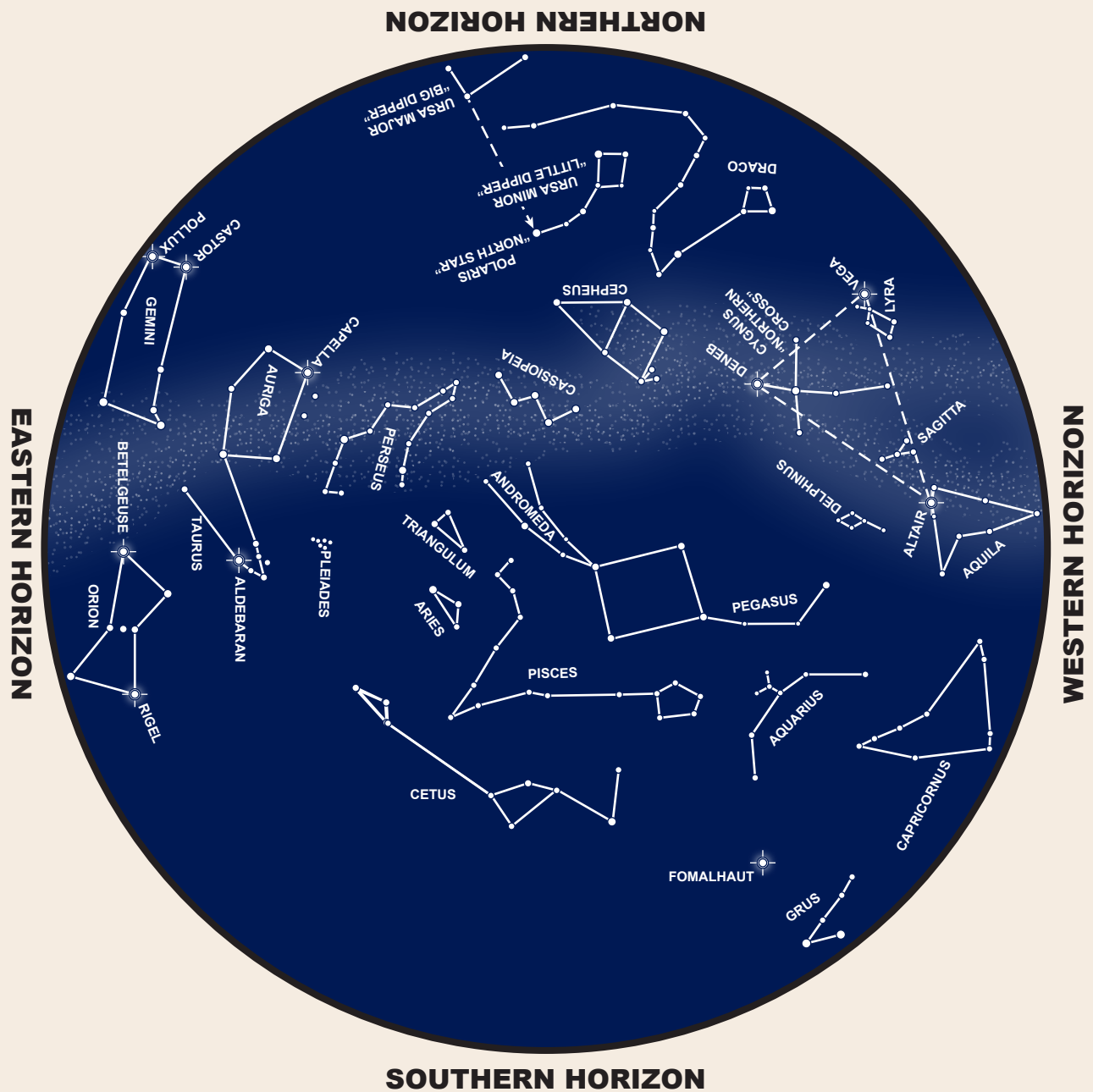
THE NIGHT SKY IN SEPTEMBER

Chart time (Daylight Saving Time):

11 pm: First of September

10 pm: Middle of September

9 pm: Last of September



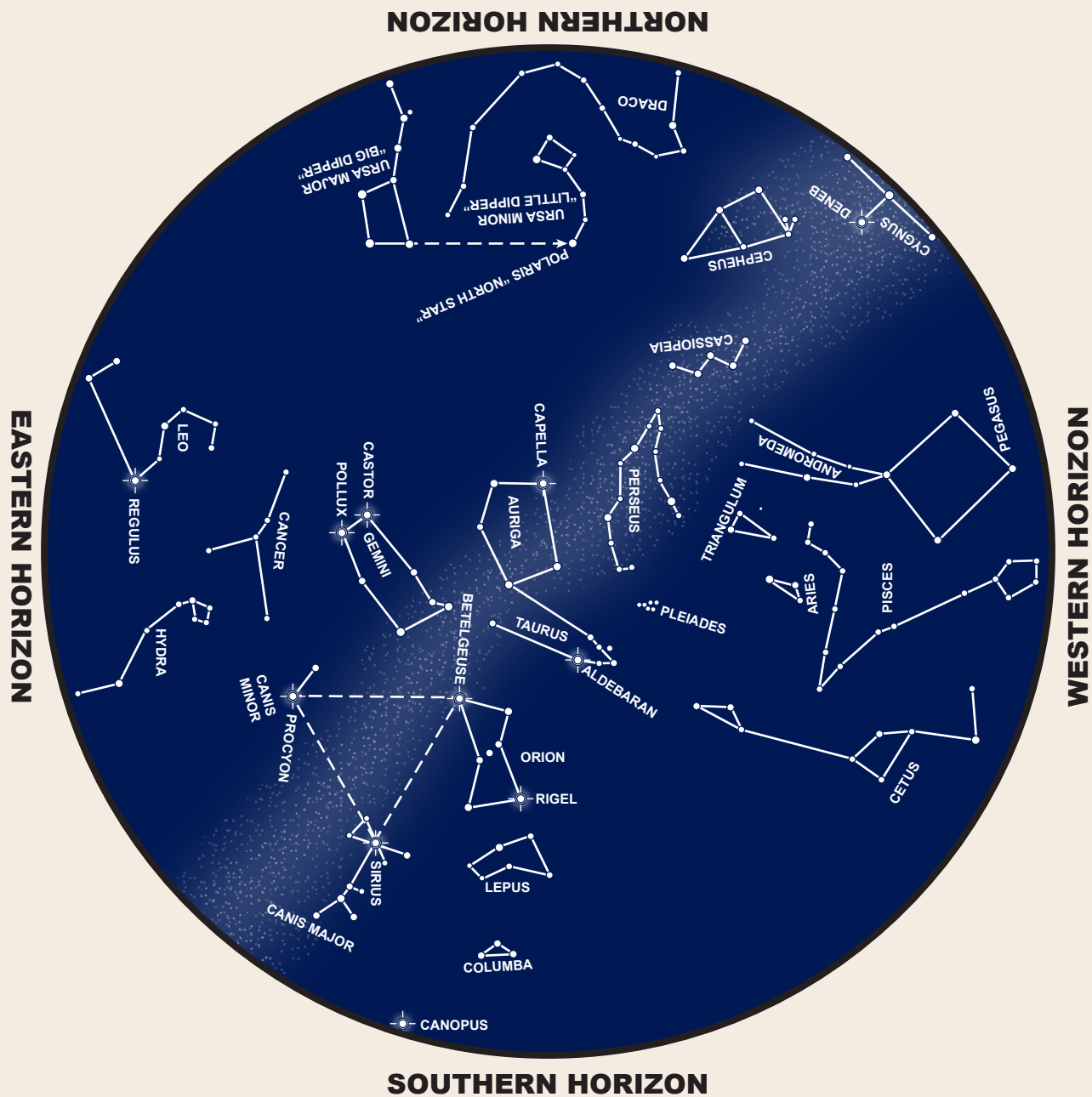
THE NIGHT SKY IN NOVEMBER

Chart time (Local Standard Time):

10 pm: First of November

9 pm: Middle of November

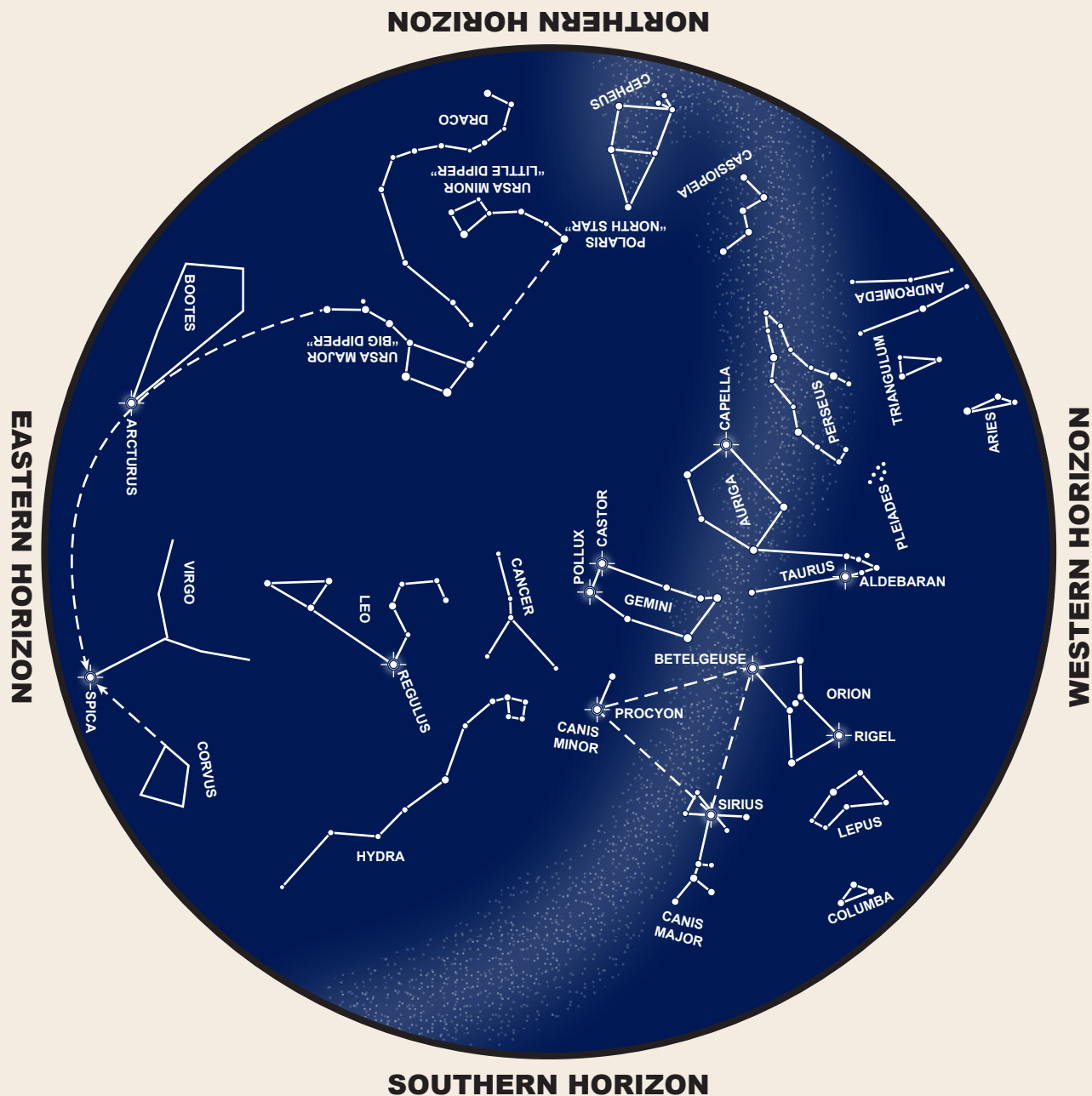
8 pm: Last of November



THE NIGHT SKY IN JANUARY

Chart time (Local Standard Time):

- 10 pm: First of January
- 9 pm: Middle of January
- 8 pm: Last of January



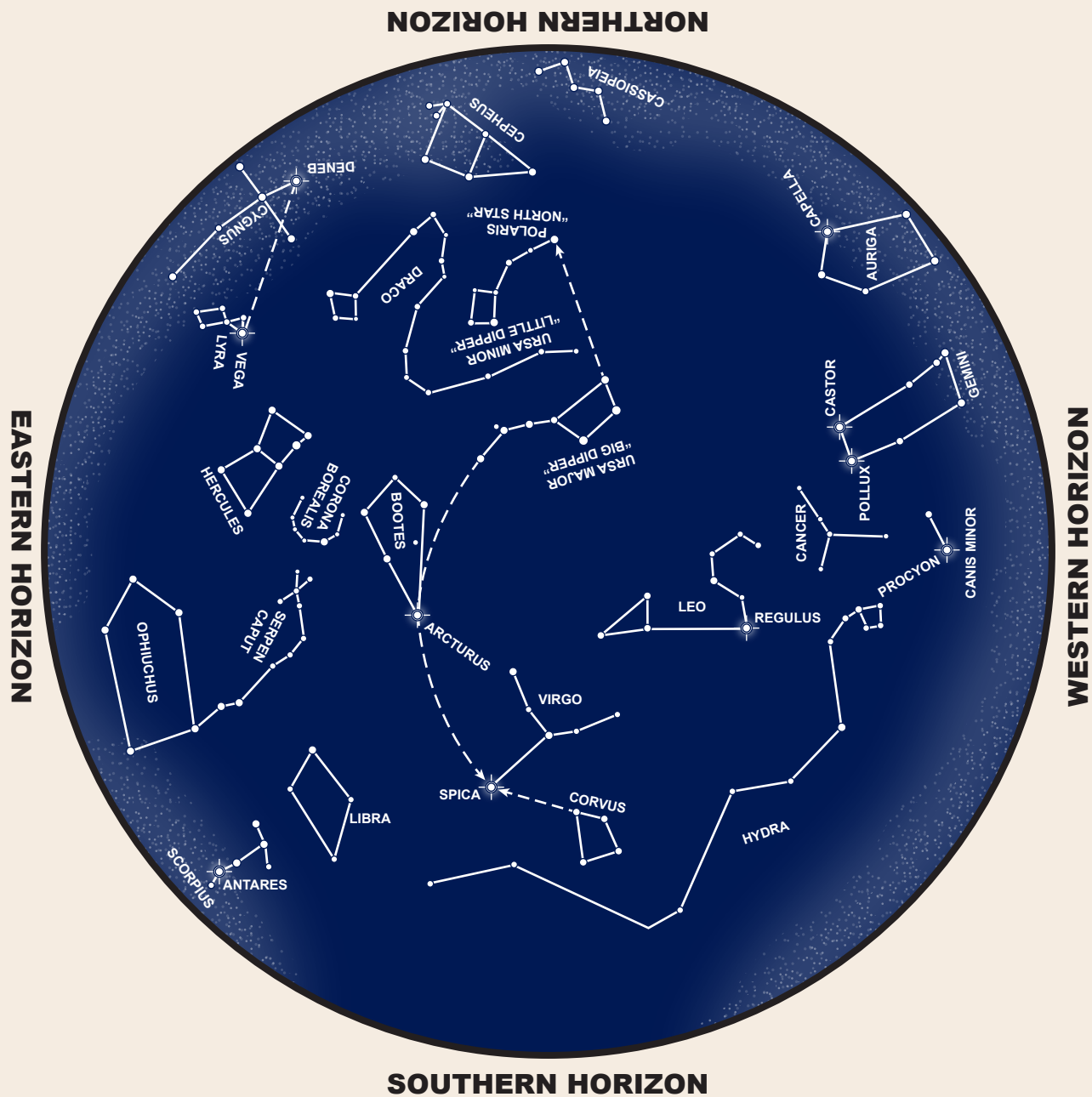
THE NIGHT SKY IN MARCH

Chart time (Local Standard Time):

10 pm: First of March

9 pm: Middle of March

8 pm: Last of March



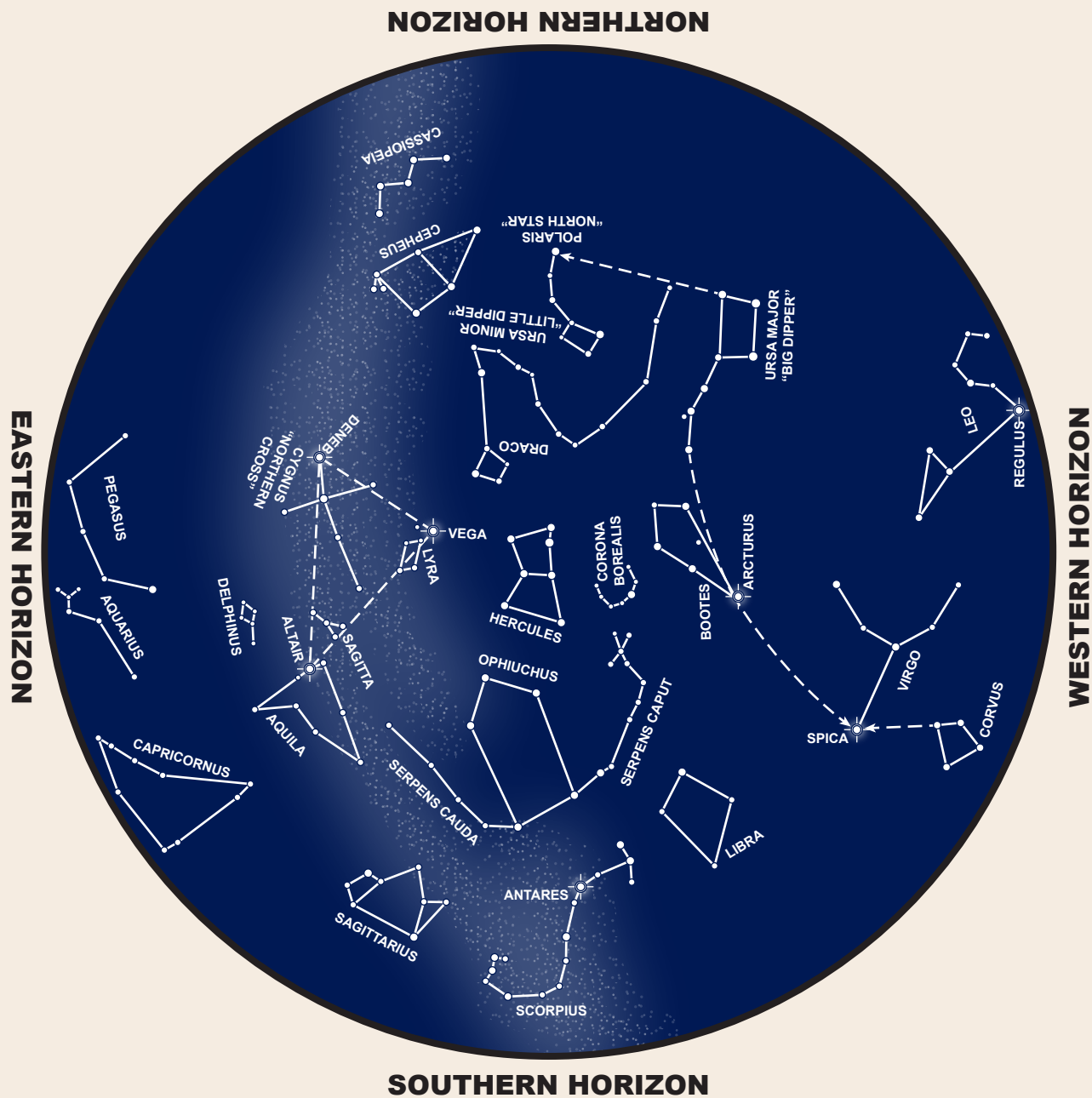
THE NIGHT SKY IN MAY

Chart time (Daylight Saving Time):

11 pm: First of May

10 pm: Middle of May

9 pm: Last of May



THE NIGHT SKY IN JULY

Chart time (Daylight Saving Time):

11 pm: First of July

10 pm: Middle of July

9 pm: Last of July

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.

ABOUT THE COVER

Meteor showers, like this glorious image of the 2012 Perseids, occur when Earth plows through streams of space debris orbiting the Sun, as discussed in Chapter 8: Vagabonds of the Solar System. Much of the debris in this image is vaporizing in our atmosphere, resulting in the distinctive streaks seen here. Each meteor shower is named after the constellation from which the meteors appear to be coming. The Perseids, for example, originate in the constellation Perseus. The precise definition of the term “constellation” is given in Chapter 1: Discovering the Night Sky and the properties of meteorites (meteors that do not vaporize completely and therefore land on Earth) are presented in Chapter 8.

